RICE UNIVERSITY

The Folding and Binding Partners of the Perlecan SEA Module

by

Ariel Diaz

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Master of Arts

APPROVED, THESIS COMMITTEE

Mary "Cindy" Farach-Carson, Professor of Biochemistry & Cell Biology

Daniel Carson, Professor of Biochemistry & Cell Biology, Dean of Natural Sciences

James McNew, Associate Professor of Biochemistry & Cell Biology, Chair

Laura Segatori, Assistant Professor of Chemical & Biomolecular Engineering

HOUSTON, TEXAS
April 2012

ABSTRACT

The Folding and Binding Partners of the Perlecan SEA Module

by

Ariel Diaz

Sperm protein, enterokinase and agrin (SEA) modules are small folds within large heavily glycosylated modular proteins. Because decreased expression of SEA-containing proteins such as perlecan (PLN) can lead to diseases such as Schwartz-Jampel syndrome (SJS), characteristics of the PLN SEA module including folding, potential for autocleavage, and protein binding were studied. Sequence analyses, recombinant protein evaluation, and a yeast two-hybrid screen were used to study the PLN SEA module and compare it to the mucin (MUC) 1 SEA module. I*n silico* modeling of the PLN SEA module demonstrated a well conserved α/β sandwich fold. Experiments with expressed proteins showed that unlike MUC1, the PLN SEA module does not autocleave. Two-hybrid screening identified four "high confidence" proteins as potential binding partners which were explored in preliminary experiments. Together, these results demonstrate that PLN SEA module is unique and its properties cannot be generalized with other SEA module proteins such as MUC1.

Acknowledgments

I would like to thank some of the many people who have helped me during the course of this project:

First I would like to thank Dr. Cindy Farach-Carson, my advisor, for all of her contributions to me professionally and personally. Her advice on research and graduate school has been wonderful. What I would most like to thank her for is her infectious optimism both in research and in life. I always felt more positive after a meeting with her.

I would also like to thank my committee for their thoughtful questions and constructive guidance on my project.

Dr. Pamela Constantinou, Brian Grindel, Patricia Chapela, and Jerahme Martinez are all greatly appreciated for their help with techniques, troubleshooting, or for directly providing research materials.

In addition to those mentioned above, I would like to thank the rest of the members of the Farach-Carson and Carson labs. They were there on good and bad research days and always helped to keep me going and smiling.

And finally to my loving husband, Michael, whose constant support and encouragement has helped me throughout this journey. Thank you for everything.

Table of Contents

List of Figures

List of Tables

Key Terms

BM	Basement membrane
DAG1	Dystroglycan
ECM	Extracellular matrix
ENTK	Enterokinase
ER	Endoplasmic reticulum
GAG	Glycosaminoglycan
HIS	Histidine
HRP	Horseradish peroxidase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ICC	Immunocytochemistry
MUC	Mucin
PDB	Protein data bank
PBS	Phosphate buffered saline solution
PLN	Perlecan
PPG	Protein Production Group
SEA	sperm protein/enterokinase/agrin modules
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis

Chapter 1

Introduction and review

SEA modules are modular folds found in highly glycosylated extracellular proteins whose functionality is poorly understood. They are small modules, consisting of approximately 120 amino acids. SEA modules are conserved in species ranging from nematodes to humans, suggesting that they have a conserved, but yet unknown, function (Bork and Patthy, 1995). Some SEA modules autocleave during cellular processing leaving the parent protein as two associated but not bound moieties (Khatri et al., 2003). The purpose of an SEA module's autocleavage is unknown but has been linked to trafficking, shedding, and compositional determination. Here I examined the PLN SEA module specifically. PLN is an extracellular matrix (ECM) protein ubiquitously expressed, that is an essential component of the basement membrane (BM) and various other territorial matrices surrounding cells (Noonan and Hassell, 1993). In instances where PLN secretion is

decreased, devastating diseases such as Dyssegmental dysplasia Silverman-Handmaker type and its milder counterpart Schwartz-Jampel syndrome (SJS) can result (OMIM, 2011). If the SEA module does indeed play a role in the trafficking of parent proteins through the secretory pathway, then a better understanding of the SEA module structure and its interactions with other proteins can provide insight into various ways of assessing diseases associated with PLN insufficiency.

1.1. Background

1.1.1. SEA modules

SEA modules were first identified in a 63-kDa sea urchin sperm protein, enterokinase (ENTK), and in another protein, agrin, giving rise to the acronym name (Bork and Patthy, 1995). SEA modules are typically 120 amino acids long and form a modular domain in glycosylated proteins. Most SEA module proteins only contain one SEA, a few contain two SEAs, and one protein, MUC16 , contains many SEA modules (Bork and Patthy, 1995; Maeda et al., 2004; Bandah-Rozenfeld et al., 2010). In membrane bound SEA-containing proteins, the module is found proximal to the cell surface.

1.1.1.1. Structure

The folding domain is composed of two primary alpha helices and three or four antiparallel beta strands that form an α/β sandwich fold or heart shaped

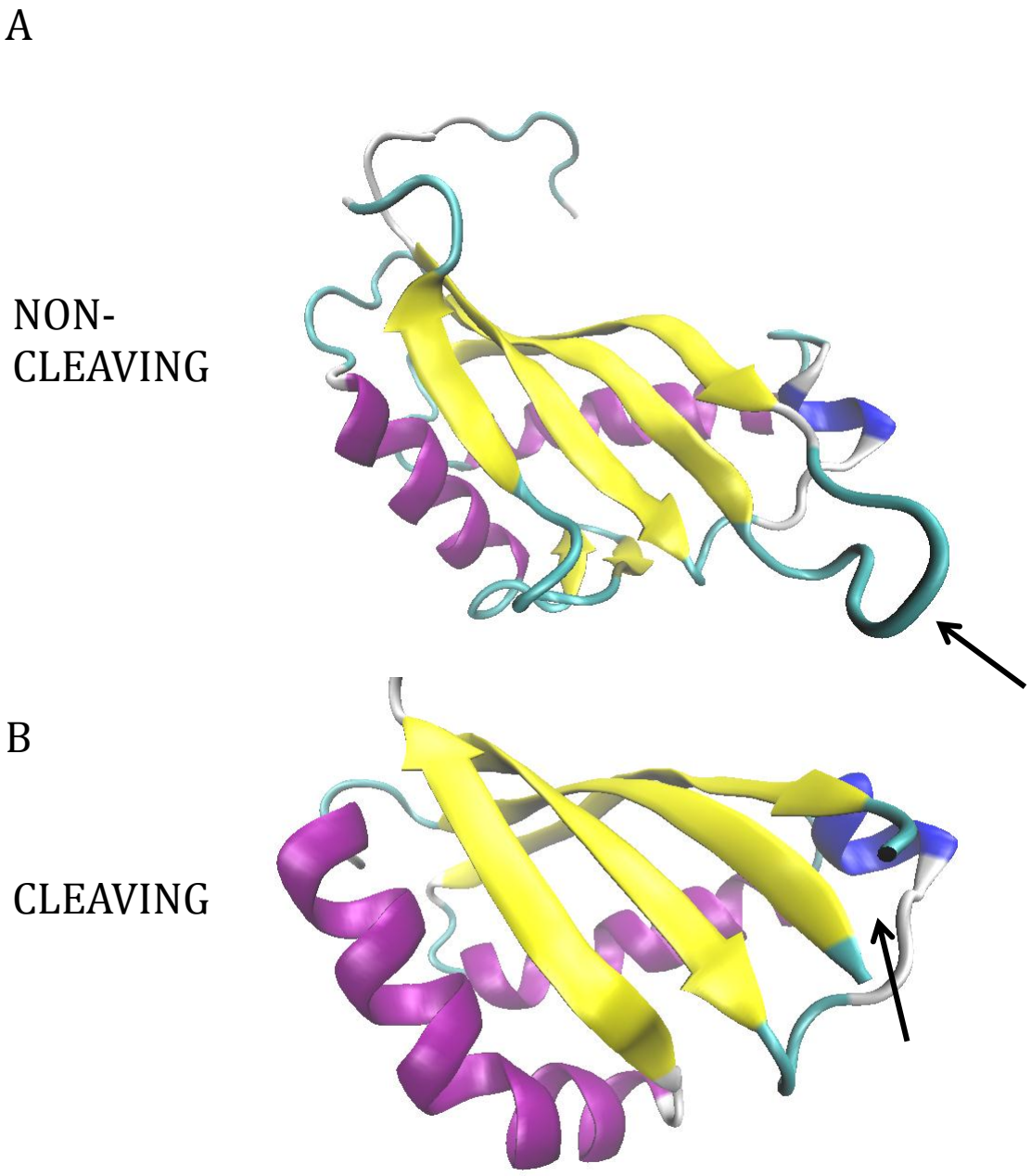


Figure 1– SEA module solution structures.
The arrow indicates the linker region between the β2 and the β3 strand. A) SEA domain from the murine MUC16 using solved using multi-dimensional NMR. The ribbon representation is of the of the lowest energy structure. The linker between β2 and the β3 strands is long and this module is not known to autocleave. B) Human MUC1 SEA module. The cleavage site is between the β2 and the β3 strand. The linker between β2 and the β3 strands is short and this module is known to autocleave. (Maeda et al., 2004 and Macao et al., 2006.)

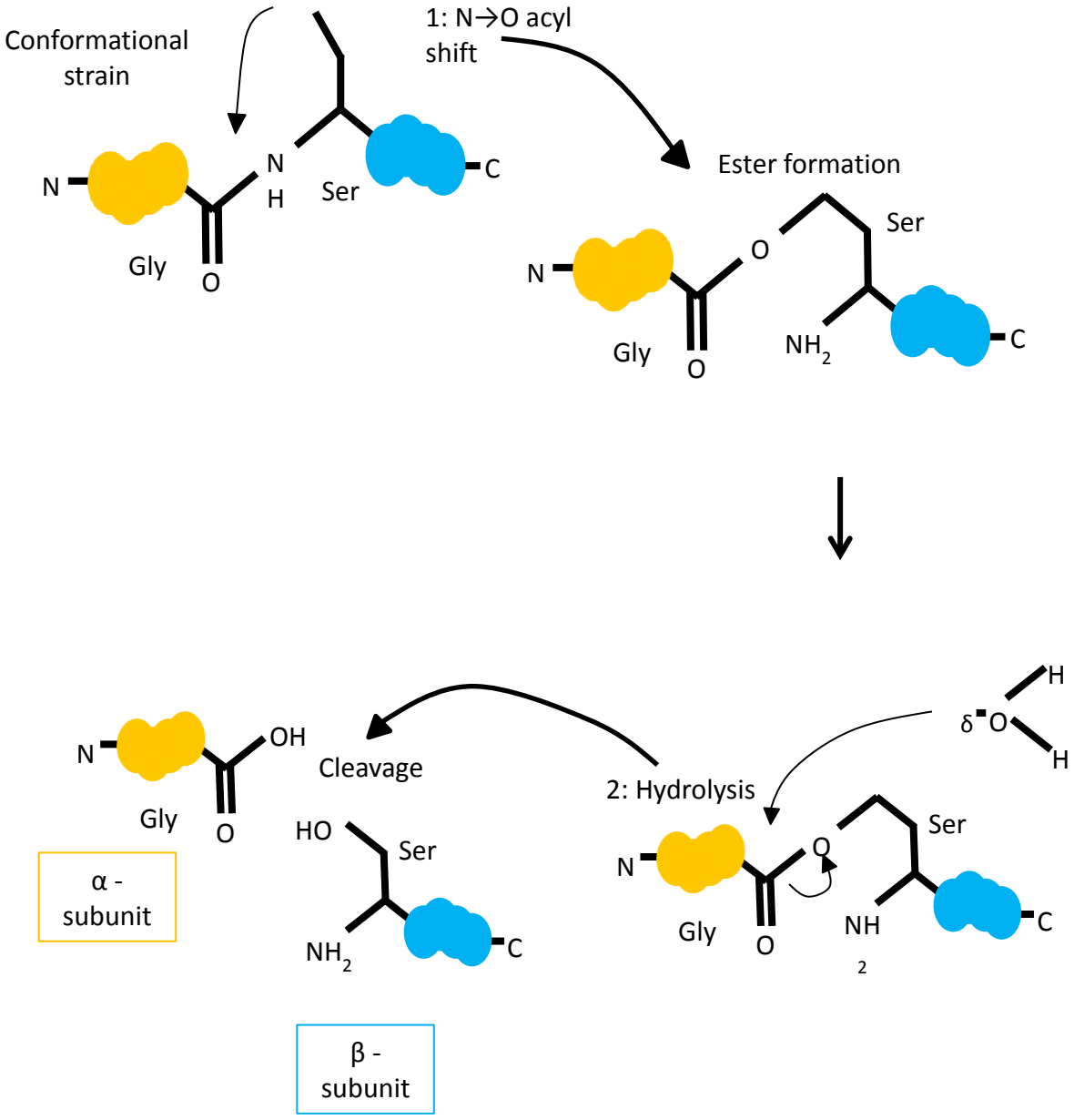


Figure 2 - Autocleavage mechanism in the MUC1 SEA.
Conformational strain on the loop between the second and third β strand from folding increases the proximity of the glycine and serine. This aides in the rearrangement reaction need to produce an ester. A hydrolysis reaction then breaks the protein into two subunits. (Levitin et al., 2005)

domain (Maeda et *al.*, 2004; Macao et *al.*, 2006). The first solution structure for an SEA module was one of the SEAs in the murine MUC16 (Protein Data Bank (PDB) ID – 1IVZ) (Figure 1) (Maeda et *al.*, 2004). The overall fold was confirmed with the solution structure of the human MUC1 SEA module (PDB ID 2ACM) (Figure 1) (Macao et *al.*, 2006). A third unpublished solution structure for the SEA of a transmembrane protease from *Mus musculus* is also in the PDB (PDB ID 2E7V).

1.1.1.2. Functional Studies

Thus far, there is no determined function of an SEA module; there are only implied ones. Some suggested functions include a role in trafficking, shedding, and compositional determination (Parry et *al.*, 2001; Wreschner et *al.*, 2002; Lillehoj et *al.*, 2003). The fact that SEA modules are found in highly glycosylated proteins suggests a potential role in mediating the passage of the protein through the secretory pathway, but this has not been demonstated. It is intriguing to suggest that SEA module interactions could facilitate the glycosylation of SEA module containing proteins either by reducing their rate of transit through the Golgi by controlling vesicle budding and progress through the secretory pathway so that carbohydrates can be added, or by stabilizing a conformation of the protein or protein complex that would increase access to the glycoslyation machinery.

The MUC1 SEA module autocleaves at a G-SVVV autocleavage motif during intracellular processing creating the two (α and β) subunits of the protein (Parry et *al.*, 2001; Levitin et *al.*, 2005; Macao et *al.*, 2006). Autocleavage in the MUC1 SEA

has been determined to be a two-step process (Levitin et al., 2005). First an N to O acyl shift, which is aided by conformational strain, creates an ester (Figure 2). This is then cleaved through a hydrolysis reaction resulting in two subunits of the protein (Figure2). This mechanism highlights the importance of the serine in the autocleavage motif. One study has shown that a threonine or a cysteine but no other amino acid in place of the serine also allows for cleavage (Macao et al., 2006). The cleavage event of the SEA module in Ig-Hepta and dystroglycan (DAG1), was found to occur in the endoplasmic reticulum (ER) and remain non-covalently associated (Abe et *al.*, 2002; Wreschner et *al.*, 2002). When the serine in the G-SVVV motif of the MUC1SEA was mutated to an alanine, it blocked cleavage and reduced the amount of shedding (Lillehoj et *al.*, 2003). Blocking cleavage also enhances the functional glycosylation of DAG1, an ECM receptor, which increases its ability to bind laminin (Wreschner et *al.*, 2002). Other data suggests that SEA modules are required for optimal biological activity in MUC3 and MUC17 (Ho et *al.*, 2010). These data point to a functional role for the SEA in the parent protein with an emphasis on the autocleavage of the SEA module being important to its function.

1.1.2. Perlecan

PLN is a modular, highly glycosylated SEA module containing heparan sulfate proteoglycan that is fully secreted into the ECM and plays an important role as a barrier, in development, and in tumorigenesis. Human PLN is 4,391 amino acids

long with a core protein that weighs approximately 450kDa. The size can be much larger depending on the amount of glycosaminoglycan (GAG) additions.

1.1.2.1. Structure

PLN has five domains (I-V) (Figure 3) (Murdoch et *al.*, 1992). Domain I contains three heparan sulfate (HS) and/or chondroitin sulfate attachment sites and an SEA module (Murdoch et al., 1992; Noonan and Hassell, 1993; Kokenyesi and Silbert, 1995; Costell et al., 1997; Dolan et al., 1997; Sasaki et al., 1998; Friedrich et al., 1999; Tapanadechopone et al., 1999). Domain I does not share homology with other domains in PLN (Murdoch et *al.*, 1992). Human PLN domain II contains four low-density lipoprotein (LDL) receptor-like modules and one immunoglobulin G (IgG)-like repeat (Murdoch et al., 1992; Noonan and Hassell, 1993). Human PLN domain III is a rigid domain that contains three laminin domain IV-like modules and eight epidermal growth factor (EGF)-like repeats. Human PLN domain IV is a large flexible domain. It is the largest of the five domains and has 21 IgG-like repeats. Human PLN domain V has three globular domains of laminin alpha-chains like modules and four EGF-like repeats (http://www.genecards.org/cgi-bin/carddisp.pl?gene=HSPG2&search=perlecan).

1.1.2.2. Function

PLN is found in bone marrow, soft tissue organs, the vasculature, and muscle tissue (Murdoch et *al.*, 1992; French et *al.*, 1999; Schofield et *al.*, 1999; Melrose et

al., 2006). It functions as a component of the BM, binds growth factors through its GAG chains, and is involved in cell signaling (Hassell et *al.*, 1980; Sasaki et *al.*, 1998; Hopf et *al.*, 1999; Kvansakul et *al.*, 2001). Perlecan helps provide an anchor for tissues and a boundary between tissues as an element of the basement membrane. It maintains the endothelial barrier as a part of the vascular ECM surrounding blood vessels. PLN is also critical to the function of the glomerular BM, a size- and charge-selective barrier. As a large multidomain protein it binds and cross-links ECM components and cell-surface molecules. It binds or interacts with nidogen-2, laminin , fibulin-2, prolargin, fibronectin, PRELP (proline/arginine-rich and leucine-rich repeat protein)/prolargin, types IV, XIII, and XVIII collagen, transthyretin and FGF-1, -2, -7, and -9 (Sasaki et *al.*, 1995; Smeland et *al.*, 1997; Ettner et *al.*, 1998; Kohfeldt et *al.*, 1998; Chang et *al.*, 2000; Mongiat et *al.*, 2000; Clayton et *al.*, 2001; Ghiselli et *al.*, 2001; Hopf et *al.*, 2001; Bengtsson et *al.*, 2002; Tu et *al.*, 2002; Salmivirta et *al.*, 2002; Miosge et *al.*, 2003). A fragment of PLN domain V called endorepellin has anti tumorigenic activity by blocking neovascularization (Mongiat et *al.*, 2003; Bix et *al.*, 2004, 2006; Fjeldstad and Kolset, 2005; Goyal et *al.*, 2011).

1.1.2.3. Pathologies

PLN acts as a boundary and when its barrier function fails, development can be compromised and tissue quality can become altered (Figure 4) and it can allow cancer cell invasion (Costell et *al.*, 1999). Homozygous mice with a null mutation

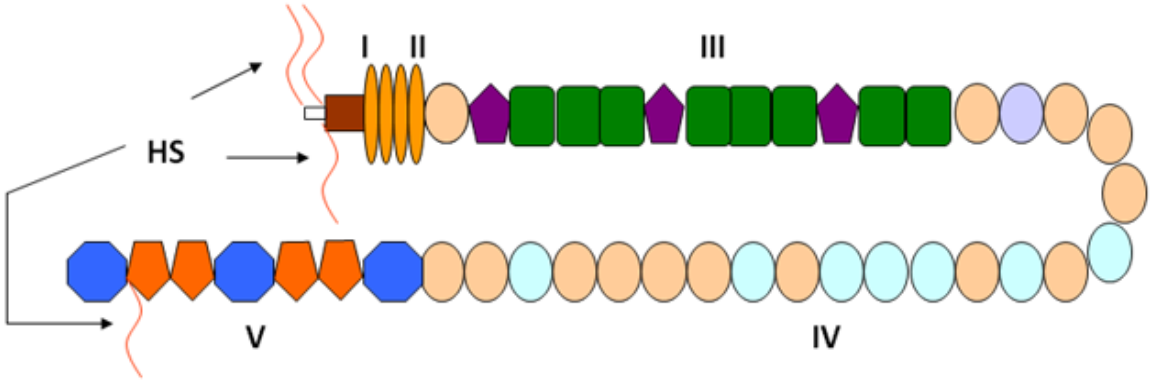


Figure 3– Perlecan schematic.
PLN has five domains,I-V. Domain I contains three HS attachment (red lines) sites and an SEA module (brown box). Human PLN domain II contains four LDL receptor-like modules and 1 IgG-like repeat. Human PLN domain III contains three laminin domain IV-like modules (purple pentagons) and eight EGF-like repeats (green squares). Domain IV has 21 IgG-like repeats . Domain V has three globular domain of laminin alpha-chains like modules (blue hexagons) and four EGF-like repeats (orange pentagons). (Modified from Farach-Carson et al., 2007)

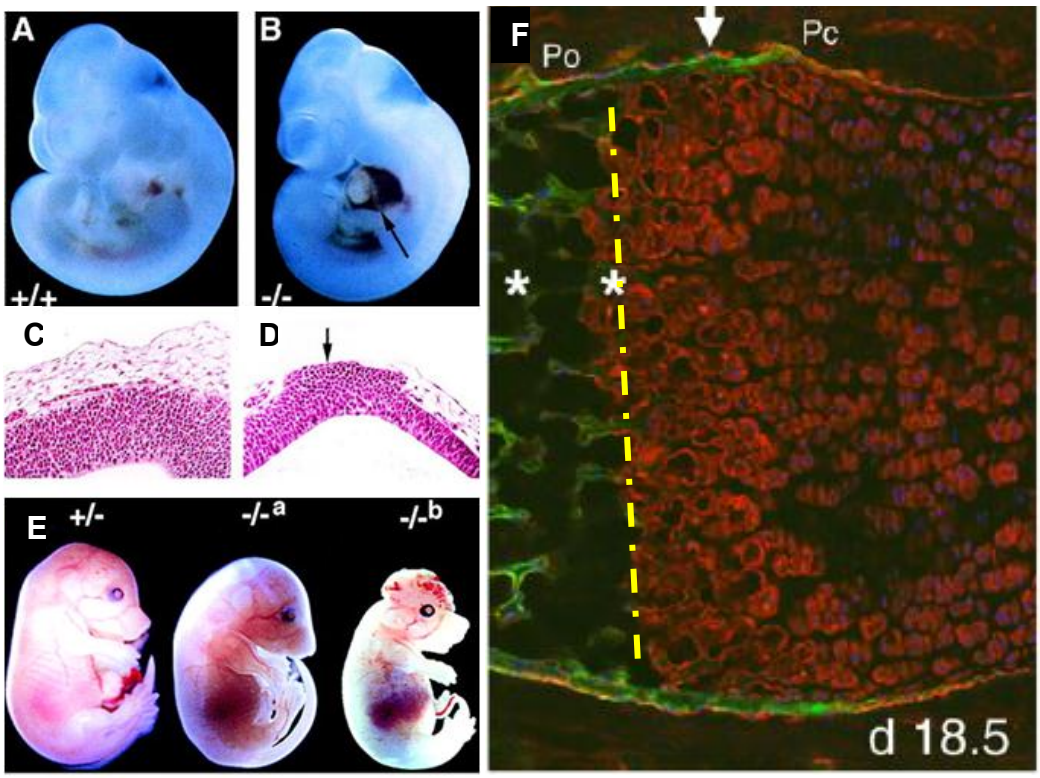


Figure 4 – Perlecan in mammals.
A) Developing mouse embryo with normal PLN expression. B) PLN-deficient embryos showing blood leakage into the pericardial cavity due to a deterioration of BMs. C) In normal brain, the neuroepithelium and the overlying mesenchyme are separated by a BM. D) Without the PLN barrier, the neuroepithelial cells invade the overlying ectoderm (arrow) which can lead to I) a loss of the roof of skull and exencephaly. C) chondro-osseous junction in a mouse hindlimb. Anti-PLN (red) in the matrix is a strong boundry between the periosteum (Po) and perichondrium (Pc). Heparanase immunostaining is green. The yellow line indicates the boundary between the developing bone and mineralized bone. (Modified from Costell et *al*., 1999 and Brown et *al*., 2008)

for PLN can create rudimentary BMs, but they deteriorate (Costell et *al.*, 1999; Sasse et *al.*, 2008). PLN is also present at a key boundary in the bone, the chondro–osseous junction (Figure 4) (Brown et *al.*, 2008). Research models with insufficient PLN secretion in the bone resemble the a chondrodysplasia disorder, SJS, characterized by reduced stature, deformed joints, pigeon breast, and other skeletal defects (Aberfeld et *al.*, 1965; Giedion, 1997; Spranger et *al.*, 2000; Rodgers et *al.*, 2007). Another example of the negative effect of the breakdown of PLN is in the kidney. When PLN is destroyed in the kidney, the filtering properties of the glomerular BM are altered and cause proteinuria (Miettinen et *al.*, 1986;Morita et *al.*, 2005). Understanding the role the PLN SEA domain plays in PLN secretion may provide insight into SJS and perhaps find new avenues for therapy.

1.1.3. Secretory Pathway and SEA proteins

A common feature of SEA module proteins is that they are glycosylated requiring them to enter the ER and proceed through the secretory pathway. The two proteins studied here have signal sequences that target them to the ER for processing and glycosylation. Human PLN's signal sequence is MGWRAAGALLLALLLHGRLLA and human MUC1's signal sequence is MTPGTQSPFFLLLLLTVLTVVTG (Parry et *al.*, 2001). The high number of hydrophobic residues in these signal sequence are typical of those that target proteins to the ER. While in the ER some SEA modules autocleave and this action

may be a check point for continued processing or an event that subsequently alters the composition of the carbohydrates added to the protein in the Golgi complex (Abe et *al.*, 2002a; Wreschner et *al.*, 2002). SEA module containing proteins are brought to the cell surface where they are either fully secreted to the extracellular environment or inserted into the membrane as highly glycosylated membrane bound extracellular proteins after successful transit through the Golgi and glycosylation.

Chapter 2

Structural aspects of the Perlecan SEA module

The PLN SEA is an 115 amino acid module found in the first domain of the protein, following the three glycosaminoglycan (GAG) chain attachment sites (Bork and Patthy, 1995). I examined structural aspects of the SEA module alone and in comparison to other known SEA modules. Sequence analyses were performed to determine consensus across a large group human SEAs as well as a narrower subset of SEA containing proteins across species. The PLN SEA was further compared to the MUC1 SEA, the most well studied SEA module. Protein folding comparisons were made with established structures and predicted structures. Similarities in parts of the sequence and structure inspired autocleavage analysis of the PLN SEA. Further sequence analysis after autocleavage studies lead to predictions of

autocleavage characteristics in unstudied proteins. Finally residues were identified that may be responsible for cleavage or binding to other proteins.

2.1. Sequence analysis of SEA modules

An initial amino acid sequence conservation analysis was performed to examine human proteins containing SEA modules using Geneious software v5.3.6 (Auckland, New Zealand). The analysis had a pairwise identity of 13.7%. There was a marked difference in the region of conservation when comparing protein sequences (Figure 5). The first half of the protein shows a higher level of conservation with little to no conserved residues in the second half of the conservation analysis.

Examining further into the amino acid conservation, I performed another analysis using a smaller subset of SEA proteins that previously had been categorized to have a shared fundamental function or a similar target molecule (Maeda et *al.*, 2004). I compared the SEA modules from PLN human and mouse; MUC3A human; MUC3B human; MUC1 mouse, hamster, gibbon, human, and cow; and ENTK pig, mouse, human, and cow. The level of conservation is higher, at 22.8% pairwise identity, than in the first analysis. This analysis maintains similar regions of conservation with the first half of the module showing a higher degree of conservation than the rest of the amino acid sequence (Figure 6). Interestingly the MUC1 SEA autocleavage motif is also very highly conserved, suggesting that the PLN SEA may also autocleave.

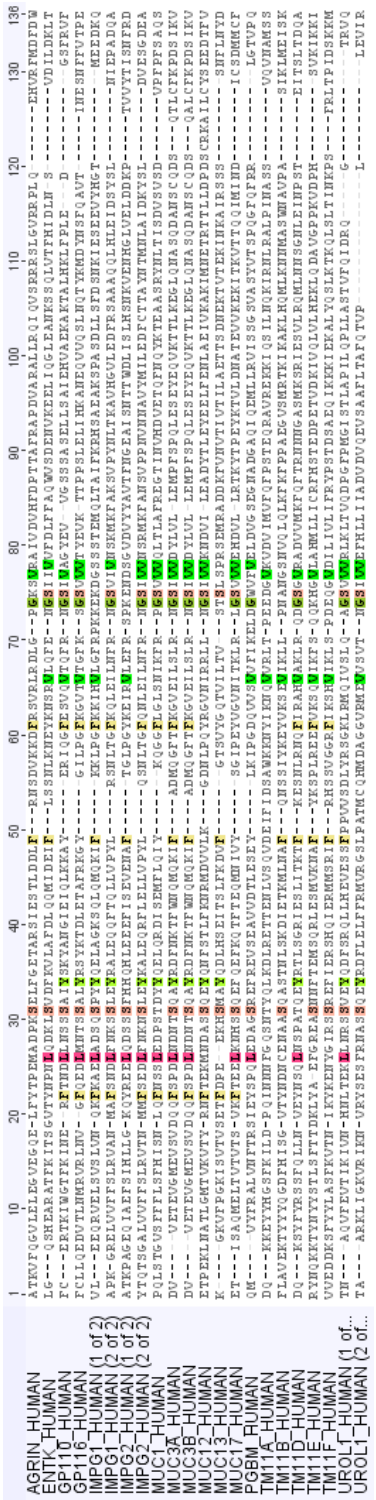


Figure 5 - Protein sequence alignment of human SEA modules.
Geneious alignment with a Blosum62 cost matrix and an exception of 1. Highlighted residues have at least a 25% consensus across the sequence alignment. SEAs in this alignment include AGRIN,ENTK, GP110, GP116, IMPG1, IMPG2, MUC1, MUC3A, MUC3B, MUC12, MUC13, PLN, TM11A, TM11B, TM11D, TM11E, TM11F, and UROL1.

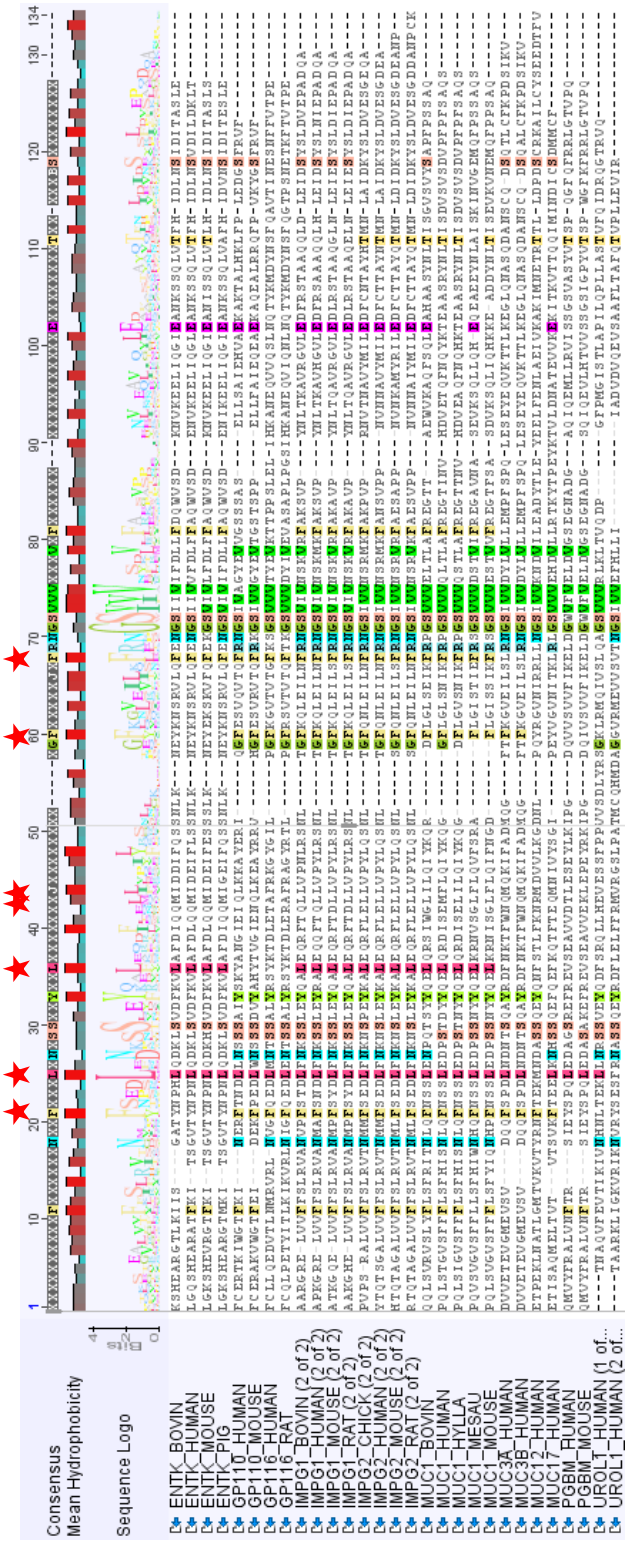


Figure 6 - Protein sequence alignment of SEA modules from similar protein families from different species.

Alignment method the same as in figure 4. Highlighted residues have at least a 25% consensus across the sequence alignment. SEAs include ENTK, GP110, GP116, IMPG1 (2 of 2), IMPG2 (2 of 2), MUC1, MUC3A, MUC3B, MUC12, MUC17, PLN and UROL1 of various species.

The MUC1 SEA module G-SVVV autocleavage motif has been highly studied. Previous studies have shown that an autocleavage event occurs between the glycine and the serine during protein processing in the ER creating two associated but non-covalently bound pieces of the SEA module (Parry et *al.*, 2001; Wreschner et *al.*, 2002; Abe et *al.*, 2002b; Khatri et *al.*, 2003; Lillehoj et *al.*, 2003; Maeda et *al.*, 2004; Levitin et *al.*, 2005; Palmai-Pallag et *al.*, 2005; Macao et *al.*, 2006). While previous studies have shown that some sequence specificity is important in this cleavage event, other have studies shown that the overall fold of the module is crucial in creating enough strain on the sequence to cause the break in the peptide back bone (Levitin et *al.*, 2005; Macao et *al.*, 2006). In figure 1 the αβ sandwich fold is maintained if both the murine MUC16 SEA and the human MUC1 sea. Not shown in the figure is the structure for the murine transmembrane protease SEA module which also folds in an αβ sandwich. The autocleavage motif in PLN is maintained with the exception of the serine and the valine. Instead its sequence reads GWVFV. This maintenance of the autocleavage motif could suggest an autocleavage event in the PLN SEA module.

2.2. Folding predictions for the Perlecan SEA module

Often the function of a protein domain is determined by its fold. In an effort to better understand the function and the role of the PLN SEA module, folding predictions were made using the Phyre2 program (Kelley and Sternberg, 2009).

Phyre2 returned 46 possible modeling templates (Table 1). The alignment guided structure prediction modeled the PLN SEA after the SEA module of the transmembrane protease 11D from *Mus musculus* (PDB ID– 2E7V) the template with the highest confidence. This folding prediction shows three beta strands and two primary alpha helices (Figure 7).

Similar folding patterns often have similar functions or characteristics and so the PLN SEA module folding prediction was compared to the more studied MUC1 SEA module that autocleaves. The fold of the MUC1 SEA module has been previously determined and maintains the α/β sandwich conformation with four antiparallel beta strands and two alpha helices (Figure 7) (Maeda et *al.*, 2004).

2.3. Characterization of recombinant SEA modules

To study the physical characteristic of the protein, I recombinantly produced it in BL21 *Escherichia coli*. Previous SEA module work has been performed with bacterially produced proteins, I wanted to avoid differences that may have arisen from glycosylation if expressed in mammalian cells (Levitin et *al.*, 2005). The MUC1 SEA module was used as a control for all experiments because it is the most well studied SEA module. Two expression constructs were made for each of the SEA modules using the inducible pET300/NT-DEST and pET301/CT-DEST destination vectors (Life technologies) that add a 6X histidine (HIS) tag to the ends of the protein for isolation and detection purposes. The protein was tagged separately at

PLN_SEA Modeling Templates

#	Template	Confidence	% i.d.	Template Information
1	c2e7vA	85.6	15	PDB header:hydrolase
				Chain: A: PDB Molecule:transmembrane protease;
				PDBTitle: crystal structure of sea domain of transmembrane protease2 from mus musculus
2	d1ivza	50.6	17	Fold:Ferredoxin-like
				Superfamily:SEA domain
				Family:SEA domain
3	c1zcoA	46.2	20	PDB header:lyase
				Chain: A: PDB Molecule:2-dehydro-3-deoxyphosphoheptonate aldolase;
				PDBTitle: crystal structure of pyrococcus furiosus 3-deoxy-d-arabino-2 heptulosonate 7-phosphate synthase
4	c2a40F	41.7	47	PDB header:structural protein
				Chain: F: PDB Molecule:wiskott-aldrich syndrome protein family member 2;
				PDBTitle: ternary complex of the wh2 domain of wave with actin-dnase i
5	c2a40C	41.7	47	PDB header:structural protein
				Chain: C: PDB Molecule:wiskott-aldrich syndrome protein family member 2;
				PDBTitle: ternary complex of the wh2 domain of wave with actin-dnase i
6	d2hgsa1	34.4	37	Fold:PreATP-grasp domain
				Superfamily:PreATP-grasp domain
				Family:Eukaryotic glutathione synthetase, substrate-binding domain
7	c3neuA	28.9	15	PDB header:structural genomics, unknown function
				Chain: A: PDB Molecule:lin1836 protein;
				PDBTitle: the crystal structure of a functionally-unknown protein lin1836 from2 listeria innocua clip11262
8	d2p02a1	22.4	24	Fold:S-adenosylmethionine synthetase
				Superfamily:S-adenosylmethionine synthetase
				Family:S-adenosylmethionine synthetase
9	c1qb3B	18.7	41	PDB header:cell cycle
				Chain: B: PDB Molecule:cyclin-dependent kinases regulatory subunit;
				PDBTitle: crystal structure of the cell cycle regulatory protein cks1
10	d1qb3a	18.5	41	Fold:Cell cycle regulatory proteins
				Superfamily:Cell cycle regulatory proteins
				Family:Cell cycle regulatory proteins

11	c2hgsA	18.4	37	PDB header:amine/carboxylate ligase
				Chain: A: PDB Molecule:protein (glutathione synthetase);
				PDBTitle: human glutathione synthetase
12	c2di3A	16.7	20	PDB header:transcription
				Chain: A: PDB Molecule:bacterial regulatory proteins, gntr family;
				PDBTitle: crystal structure of the transcriptional factor cgl29152 from corynebacterium glutamicum
13	d3c7bb2	13	18	Fold:Ferredoxin-like
				Superfamily:Nitrite/Sulfite reductase N-terminal domain-like
				Family:DsrA/DsrB N-terminal-domain-like
14	d1mxaa1	12.9	23	Fold:S-adenosylmethionine synthetase
				Superfamily:S-adenosylmethionine synthetase
				Family:S-adenosylmethionine synthetase
15	c2wfoA	12.6	24	PDB header:viral protein
				Chain: A: PDB Molecule:glycoprotein 1;
				PDBTitle: crystal structure of machupo virus envelope glycoprotein2 gp1
16	d1qm4a1	11.5	21	Fold:S-adenosylmethionine synthetase
				Superfamily:S-adenosylmethionine synthetase
				Family:S-adenosylmethionine synthetase
17	d1m0wa1	11.5	35	Fold:PreATP-grasp domain
				Superfamily:PreATP-grasp domain
				Family:Eukaryotic glutathione synthetase, substrate-binding domain
18	d2vv5a2	10.7	27	Fold:Ferredoxin-like
				Superfamily:Mechanosensitive channel protein MscS (YggB), C-terminal domain
				Family:Mechanosensitive channel protein MscS (YggB), C-terminal domain
19	c2ky5A	10.3	43	PDB header:cell adhesion
				Chain: A: PDB Molecule:platelet endothelial cell adhesion molecule;
				PDBTitle: solution structure of the pecam-1 cytoplasmic tail with dpc
20	c3kalB	10.1	35	PDB header:ligase
				Chain: B: PDB Molecule:homoglutathione synthetase;
				PDBTitle: structure of homoglutathione synthetase from glycine max in2 closed conformation with homoglutathione, adp, a sulfate3 ion, and three magnesium ions bound
21	c1y7xA	9.8	21	PDB header:structural protein, protein binding
				Chain: A: PDB Molecule:major vault protein;
				PDBTitle: solution structure of a two-repeat fragment of major vault2 protein

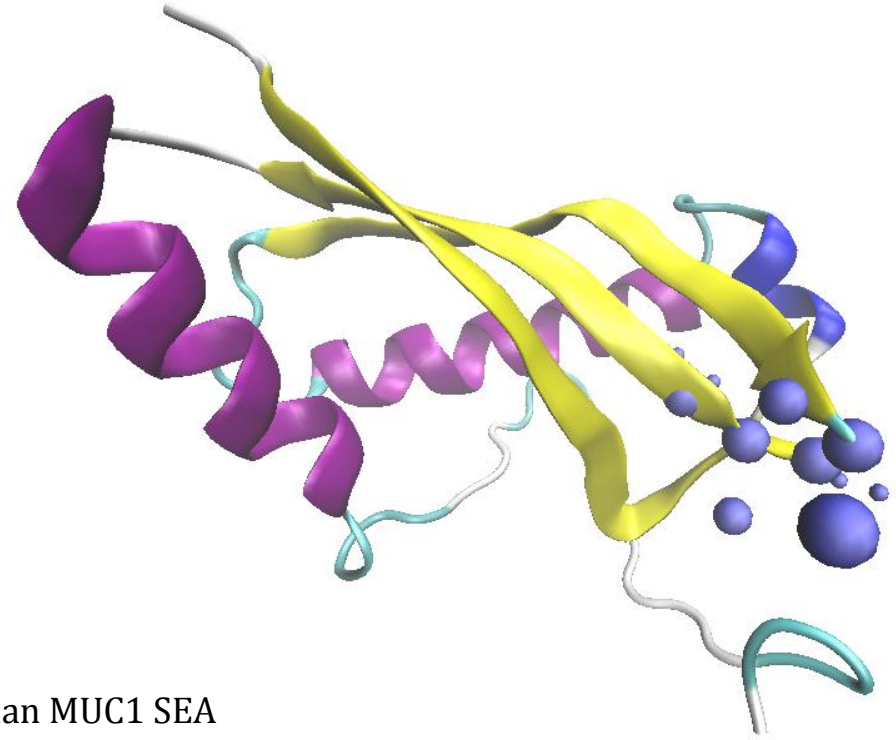
22	c1xbsA	9.7	20	PDB header:transcription, cell cycle Chain: A: PDB Molecule:dim1-like protein; PDBTitle: crystal structure of human dim2: a dim1-like protein
23	c3t4cD	9.3	12	PDB header:transferase Chain: D: PDB Molecule:2-dehydro-3-deoxyphosphooctonate aldolase 1; PDBTitle: crystal structure of 2-dehydro-3-deoxyphosphooctonate aldolase from2 burkholderia ambifaria
24	d1r26a	9.2	14	Fold:Thioredoxin fold Superfamily:Thioredoxin-like Family:Thioltransferase
25	c1m0tB	8.4	35	PDB header:ligase Chain: B: PDB Molecule:glutathione synthetase; PDBTitle: yeast glutathione synthase
26	c3ic7A	8	7	PDB header:structural genomics, unknown function Chain: A: PDB Molecule:putative transcriptional regulator; PDBTitle: crystal structure of putative transcriptional regulator of gntr family2 from bacteroides thetaiotaomicron
27	c2y75F	7.7	15	PDB header:transcription Chain: F: PDB Molecule:hth-type transcriptional regulator cymr; PDBTitle: the structure of cymr (yrzc) the global cysteine regulator2 of b. subtilis
28	c3kfwX	7.5	10	PDB header:structural genomics, unknown function Chain: X: PDB Molecule:uncharacterized protein; PDBTitle: uncharacterized protein rv0674 from mycobacterium tuberculosis
29	c1cffB	7.4	50	PDB header:calmodulin Chain: B: PDB Molecule:protein (calcium pump); PDBTitle: nmr solution structure of a complex of calmodulin with a2 binding peptide of the ca2+-pump
30	c3eetA	7.4	16	PDB header:transcription regulator Chain: A: PDB Molecule:putative gntr-family transcriptional regulator; PDBTitle: crystal structure of putative gntr-family transcriptional2 regulator
31	c2i7fB	7.1	20	PDB header:oxidoreductase Chain: B: PDB Molecule:ferredoxin component of dioxygenase; PDBTitle: sphingomonas yanoikuyae b1 ferredoxin

32	c3fhkF	7.1	19	PDB header:structural genomics, unknown function
				Chain: F: PDB Molecule:upf0403 protein yphp;
				PDBTitle: crystal structure of apc1446, b.subtilis yphp disulfide2 isomerase
33	d2r7ca2	7	27	Fold:Rotavirus NSP2 fragment, N-terminal domain
				Superfamily:Rotavirus NSP2 fragment, N-terminal domain
				Family:Rotavirus NSP2 fragment, N-terminal domain
34	c1g5vA	6.6	17	PDB header:translation
				Chain: A: PDB Molecule:survival motor neuron protein 1;
				PDBTitle: solution structure of the tudor domain of the human smn2 protein
35	d1lvaa3	6.6	23	Fold:DNA/RNA-binding 3-helical bundle
				Superfamily:"Winged helix" DNA-binding domain
				Family:C-terminal fragment of elongation factor SelB
36	c1trlB	6.4	25	PDB header:hydrolase (metalloprotease)
				Chain: B: PDB Molecule:thermolysin fragment 255 - 316;
				PDBTitle: nmr solution structure of the c-terminal fragment 255-3162 of thermolysin: a dimer formed by subunits having the3 native structure
37	c2gu0A	6.4	33	PDB header:viral protein
				Chain: A: PDB Molecule:nonstructural protein 2;
				PDBTitle: crystal structure of human rotavirus nsp2 (group c /2 bristol strain)
38	c3pnwX	6.2	50	PDB header:protein binding/immune system
				Chain: X: PDB Molecule:tudor domain-containing protein 3;
				PDBTitle: crystal structure of the tudor domain of human tdrd3 in complex with2 an anti-tdrd3 fab
39	c2d9tA	6	50	PDB header:structural genomics, unknown function
				Chain: A: PDB Molecule:tudor domain-containing protein 3;
				PDBTitle: solution structure of the tudor domain of tudor domain2 containing protein 3 from mouse
40	d1mhna	6	17	Fold:SH3-like barrel
				Superfamily:Tudor/PWWP/MBT
				Family:Tudor domain

41	d1fe0a	6	13	Fold:Ferredoxin-like
				Superfamily:HMA, heavy metal-associated domain
				Family:HMA, heavy metal-associated domain
42	c1vs1B	5.7	26	PDB header:transferase
				Chain: B: PDB Molecule:3-deoxy-7-phosphoheptulonate synthase;
				PDBTitle: crystal structure of 3-deoxy-d-arabino-heptulosonate-7-2 phosphate synthase (dahp synthase) from aeropyrum pernix3 in complex with mn2+ and pep
43	d1wy6a1	5.7	21	Fold:alpha-alpha superhelix
				Superfamily:Hypothetical protein ST1625
				Family:Hypothetical protein ST1625
44	c4a4fA	5.6	17	PDB header:rna binding protein
				Chain: A: PDB Molecule:survival of motor neuron-related-splicing factor 30;
				PDBTitle: solution structure of spf30 tudor domain in complex with2 symmetrically dimethylated arginine
45	d1ldda	5.3	16	Fold:DNA/RNA-binding 3-helical bundle
				Superfamily:"Winged helix" DNA-binding domain
				Family:SCF ubiquitin ligase complex WHB domain
46	d1knwa1	5.2	30	Fold:Domain of alpha and beta subunits of F1 ATP synthase-like
				Superfamily:Alanine racemase C-terminal domain-like
				Family:Eukaryotic ODC-like

Table 1 – Phyre2 results for the PLN SEA module.
Phyre2 returned proteins that could be used as a template for modeling the PLN SEA module. Each protein is listed by its PDB id, given a confidence score, the percent of matching protein sequence is listed, and some descriptive information.

A Human PLN SEA



B Human MUC1 SEA

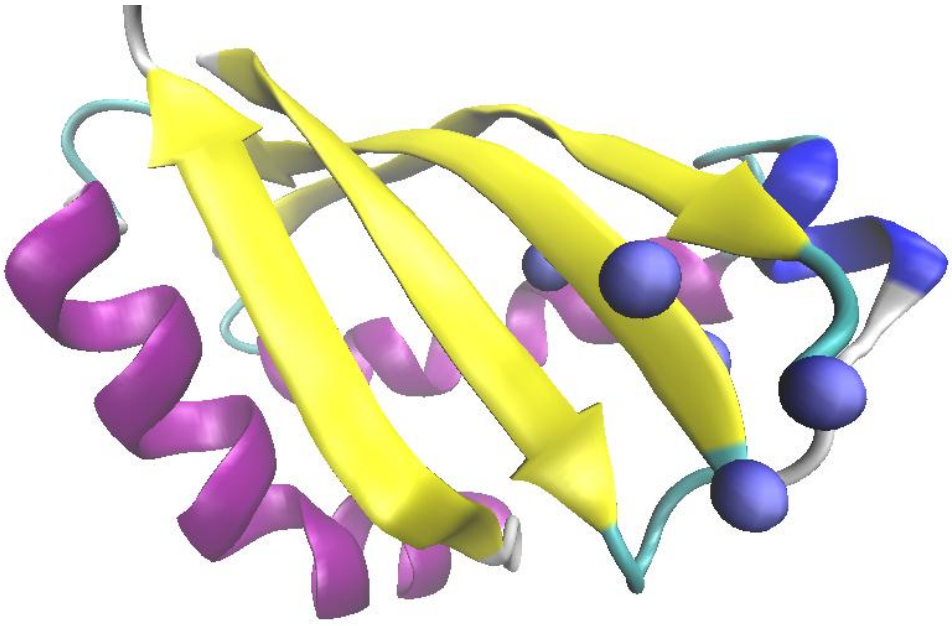


Figure 7 – SEA module folding.
Structures are displayed in the NewCartoon style with VMD software. Individual residues are highlighted as beads. A) Predicted folding of PLN SEA module by Phyre2 base on PDB structure 2E7V. Residues 62-67 (GWVFV) of the SEA are highlighted, corresponding to residues 142 – 146 of PLN. B) MUC1 SEA folf (PDB 2ACM). Residues 64 – 68 (GSVVV) are highlighted. These correspond to residues 1097 – 1011 of MUC1.

either end so that both fragments could be detected in the event of autocleavage. The protein of interest was extracted from cell lysates by metal ion chromatography. Samples from both the whole cell lysate and purified protein were analyzed.

Protein production was confirmed by Coomassie staining of pre and post induction samples of whole cell lysates on sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gels (Figure 8). The arrows indicate new protein bands in the post induction sample not found in the pre induction sample. The strength of the post induction bands indicates very high abundance of recombinant protein produced by the cultures. Purified protein was then dialyzed into elution buffers with pHs ranging from five to eight to examine the stability of the recombinant protein. There was no change from sample to sample across all the pHs tested (Figure 9). None of the samples showed new bands at smaller than expected sizes that would indicate a pH dependent autocleavage.

2.3.1. Perlecan SEA module autocleavage analysis

SEA module autocleavage analysis was initially examined by Coomassie staining of whole cell lysates on gradient SDS-PAGE gels (Figure 8). The new band in the post induction sample for the N-terminally tagged PLN (PN) SEA module is approximately 14kDa, the same size as the predicted weight of the full length PLN SEA. The results were not repeated for the C-terminally tagged PLN (PC) SEA. The new band in the post induction sample was approximately 28kDa, which is neither

the size of the full length or the cleaved protein. This band is twice the size of the expected protein. Sequencing the expression plasmid to verify that only one copy of the SEA module is present would eliminate one possible reason for the presence of this 28kDa band. Similarly inconclusive results were seen in the expressed MUC1 SEA-module samples (Data not shown).

To obtain more conclusive autocleavage information, western blots were performed on the purified proteins (Figure 10). The PLN SEA was detected at approximately 14kDa, similar to the initial Coomassie result and matching the expected size of the full length protein. If the PLN SEA were cleaved a band would have been detected at about half the size. The MUC1 SEA was detected at approximately 7kDa. This size fragment matches results from previous studies showing that the MUC1 SEA autocleaves approximately in the middle of the sequence at the G-SVVV autocleavage motif (Levitin et *al.*, 2005). The upper three bands are likely bacterial protein because they are found in both samples. To clarify this, the antibody should be tested on bacterial whole cell lysate to see if the same bands are found in uninduced unpurified protein samples.

Models of small pieces of each protein containing the autocleavage motif were made to better understand the differences in autocleavage results. The Phyre2 program (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) was used and returned a number of possible templates for each protein (Table 2 and 3). Only the template with the highest confidence was used to create the model. The MUC1

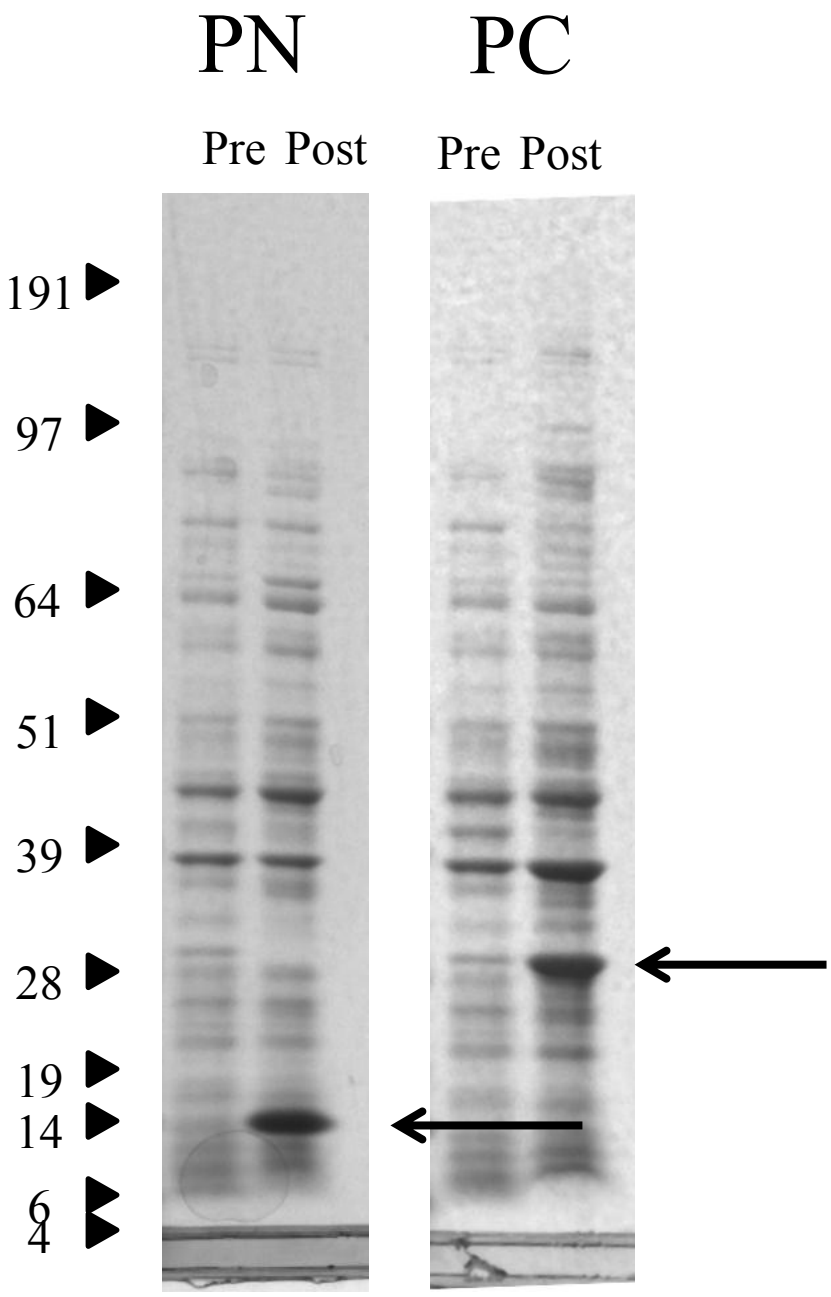


Figure 8 - Coomassie staining of pre and post induction samples.
Whole cell lysates from recombinant protein producing bacterial cultures were electrophoresed on gradient SDS-PAGE gels. This gel both confirms production and indicates the size of the recombinant proteins. PN recombinant protein in the post induction sample is detected as a 14kDa band. PC recombinant protein is detected as a 28kDa band in the post induction sample. The abundance of both recombinant protein is very high.

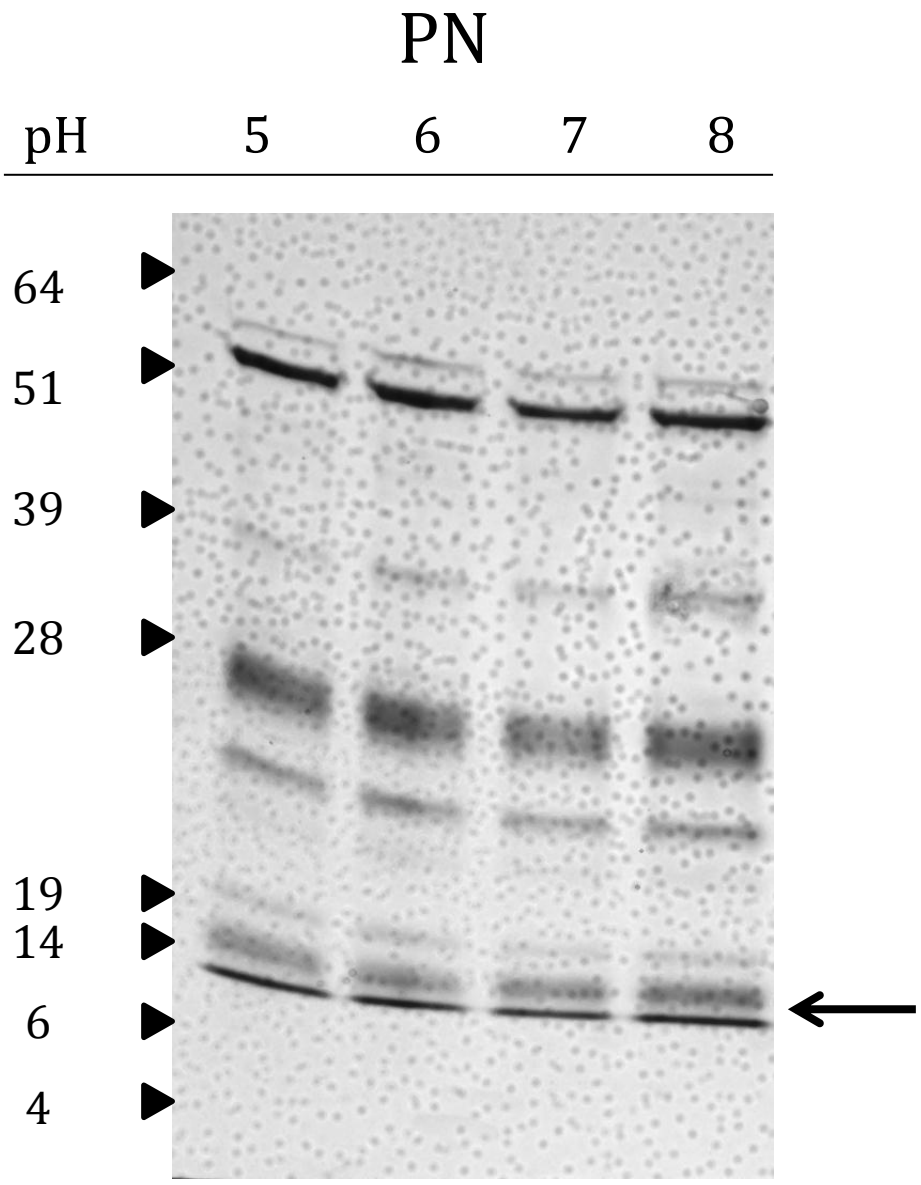


Figure 9 – pH stability of recombinant protein PN.
Purified protein PN was dialyzed into elution buffers (50mM NaH_2PO_4, 300mM NaCl and 300mM imidazole) of different pHs and maintained for 16 hours at 4°C. Samples of each were electrophoresed on gradient SDS-PAGE gels and Coomassie stained. The arrow indicates the expected protein size. No band is detected at a smaller then expected size in any of the samples.

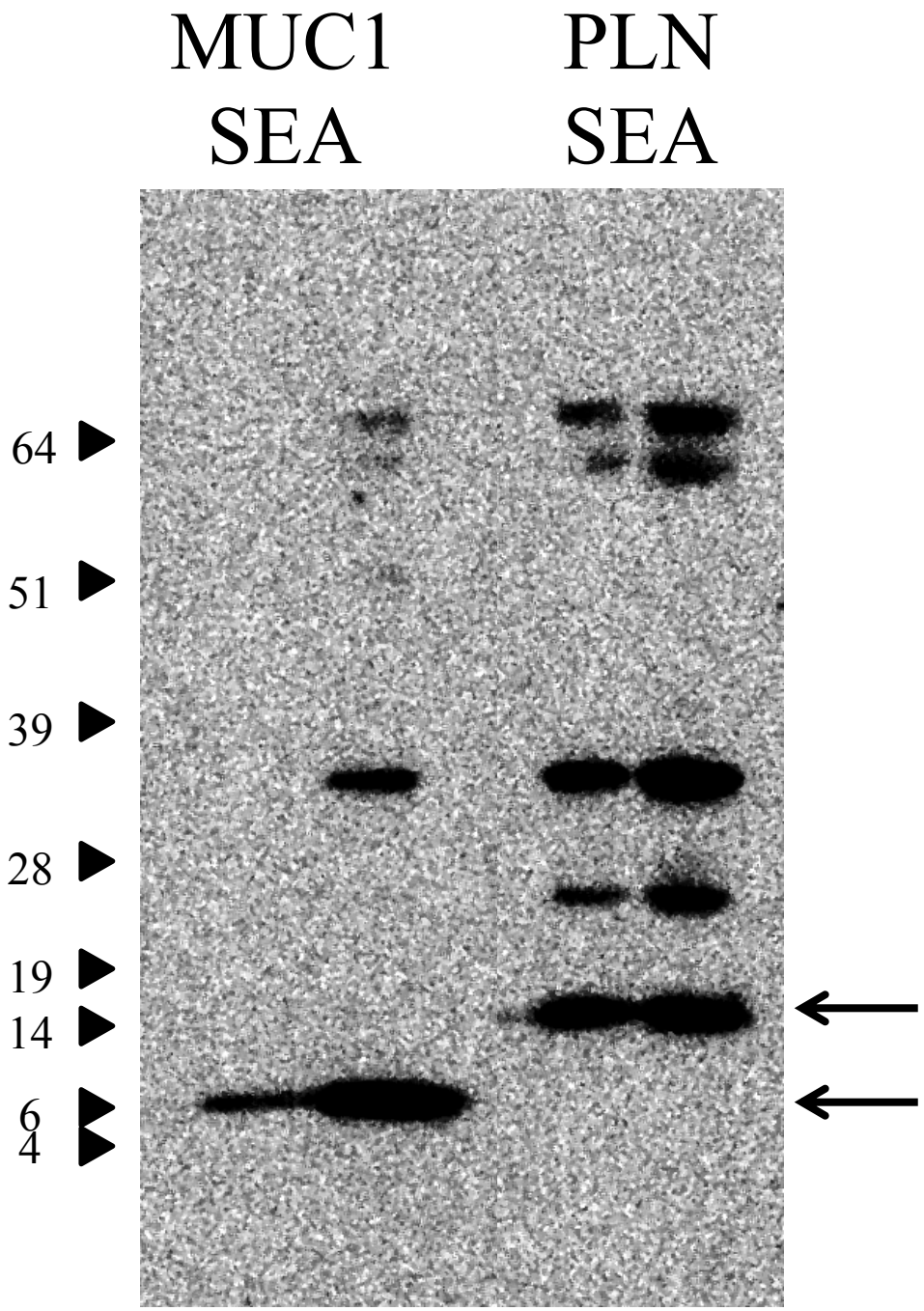


Figure 10 – Western blot autocleavage analysis.
Purified recombinant proteins were electrophoresed on a NuPAGE® Novex® 4-12% Bis-Tris gradient SDS-PAGE gel in MOPS running buffer. The primary antibody was an anti-HIS mouse monoclonal antibody(Qiagen) and the secondary antibody wand and anti-mouse HRP antibody (BioRad). MUC1 SEA recombinant protein detected as a 6kDa band. PLN SEA recombinant protein is detected as a 14kDa band. Larger protein bands are likely bacterial proteins as they are detected in both samples.

structure (PDB ID – 2ACM) was used in the modeling of the sequence surrounding and including the autocleavage motif for MUC1. The confidence in this model was high at 97.2. The MUC1 model shows a tight turn just before the autocleavage motif between the two beta strands (Figure 11). This tight turn is what has been shown to be necessary to place enough steric strain on the peptide backbone to cause autocleavage (Macao et *al.*, 2006). The PLN autocleavage motif was model using the putative transcription anti-termination protein NusG (PDB ID – c3ewgA) as the template. The confidence in this model is low at 28.3. The PLN SEA model shows a much longer loop between the two strands and preceding the GWVFV sequence (Figure 11).

2.4. Autocleavage predictions

In an attempt to make autocleavage predictions on unstudied proteins a sequence analysis was performed on the SEA modules that are known to autocleave; DAG1, G-protein coupled receptor 116 (GP116), MUC1, MUC3A, MUC3B, MUC12, and TMPRSS6. A sequence analysis was performed and only the residues that were 100% conserved were noted (Figure 12). The analysis showed a 24.8% matching sequencing identity. Three residues were identified from the alignment; F22, G70, and S71. The phenylalanine is not associated with any region of the SEA of interest thus far. The glycine and serine are those of the autocleavage motif. SEA sequences that are known to autocleave were analyzed by Phyre2 to see where the phenylalanine identified by sequence alignment is located in the folded protein. The

phenylalanine is located next to the glycine and the serine of the autocleave region (Figure 13). While the phenylalanine would not seem to directly participate in the reactions depicted in figure 2, it may press on the autocleavage motif, increasing the strain on the loop between the β2 and β3 strands, indirectly aiding in autocleavage. Using the criteria of having a phenylalanine that aligns with the one identified in figure 12 and having both the glycine and the serine of the autocleavage motif predictions were made as to which SEA will autocleave and which ones will not (Table 4).

2.5. Identification of residues potentially involved in SEA binding

I chose to focus on the hydrophobic residues and the potential for binding pocket on the SEA module. In collaboration with another graduate student at Rice University, Drew Bryant (Department of Computer Science), a ClustalW sequence alignment and identified a number of highly conserved residues (Data not shown) (Thompson et *al.*, 1994). These residues are noted on my alignment in figure 6. The residues identified were mapped to a space filling model of the MUC1 SEA (Figure 14). The residues were reduced to only those that are solvent facing. The residues identified were F21, L25, L36, L43, L44, F60, and F68 of the sequence alignment and consensus sequence. These residue locations map to Y17, L21, V32, T39, L40, V52, and E50 of the PLN SEA module sequence. The residues cluster in two different

areas of the SEA module and are not evenly distributed. One of the hydrophobic pockets is located near the MUC1 autocleavage site as noted in the 180° view.

2.6. Materials and Methods

2.6.1. Sequence analysis

Proteins with accessions numbers, O00468, P98072, P98073, P97435, P98074, Q5T601, Q8VEC3, Q8IZF2, Q9WVT0, Q9GMS5, Q17R60, Q8R1W8, Q9ET62, Q1XI86, Q9BZV3, Q80XH2, P70628, Q8WML4, P15941, Q29435, Q60528, Q02496, Q02505, Q9H195, Q9UKN1, Q9H3R2, Q685J3, P98160, Q05793, Q6ZMR5, Q86T26, O60235, Q9UL52, Q6ZWK6, Q5DID0, were downloaded using the Uniprot search function in Geneious software v5.3.6. More information on each of the proteins can be found in table five.

The SEA module sequence from each of the proteins listed above were extracted from the parent protein sequence and used in sequence alignments. Information on each of the SEA module sequences can be found in table six. The first alignment compared the sequences of human SEA modules from agrin, ENTK, GP110, GP116, IMPG1, IMPG2, MUC1, MUC3A, MUC3B, PLN, TM11A, TM11B, TM11D, TM11E, TM11F, and UROL1 using a Geneious alignment. The alignment was a global alignment without free end gaps. It used a Blosum62 (BLOcks of Amino Acid SUbstitution Matrix) cost matrix, an open gap penalty of 12, a gap extension penalty of three, and an exception of one. The alignment was run through ten

PLN_SEA_GWVFV Modeling Templates

#	Template	Confidence	% i.d.	Template Information
1	c3ewgA_	28.3	35	PDB header:transcription
				Chain: A: PDB Molecule:putative transcription antitermination protein nusg;
				PDBTitle: crystal structure of the n-terminal domain of nusg (ngn) from2 methanocaldococcus jannaschii
2	c3h7hB_	21.4	35	PDB header:transcription
				Chain: B: PDB Molecule:transcription elongation factor spt5;
				PDBTitle: crystal structure of the human transcription elongation factor dsif,2 hspt4/hspt5 (176-273)
3	c2exuA_	19.4	17	PDB header:transcription
				Chain: A: PDB Molecule:transcription initiation protein spt4/spt5;
				PDBTitle: crystal structure of saccharomyces cerevisiae transcription elongation2 factors spt4-spt5ngn domain
4	c3p8bB_	18.9	20	PDB header:transferase/transcription
				Chain: B: PDB Molecule:transcription antitermination protein nusg;
				PDBTitle: x-ray crystal structure of pyrococcus furiosus transcription2 elongation factor spt4/5
5	d1hh2p1	14.6	17	Fold:OB-fold
				Superfamily:Nucleic acid-binding proteins
				Family:Cold shock DNA-binding domain-like
6	c2k9kA_	10.9	17	PDB header:metal transport
				Chain: A: PDB Molecule:tonb2;
				PDBTitle: molecular characterization of the tonb2 protein from vibrio2 anguillarum
7	d2gskb1	10.3	38	Fold:TolA/TonB C-terminal domain
				Superfamily:TolA/TonB C-terminal domain
				Family:TonB
8	d1ihra_	9.6	38	Fold:TolA/TonB C-terminal domain
				Superfamily:TolA/TonB C-terminal domain
				Family:TonB
9	c3d89A_	9.4	24	PDB header:electron transport
				Chain: A: PDB Molecule:rieske domain-containing protein;
				PDBTitle: crystal structure of a soluble rieske ferredoxin from mus musculus
10	c2x3wD_	8.4	24	PDB header:endocytosis
				Chain: D: PDB Molecule:protein kinase c and casein kinase substrate in neurons
				PDBTitle: structure of mouse syndapin i (crystal form 2)

11	d2cp5a1	7.1	38	Fold:SH3-like barrel
				Superfamily:Cap-Gly domain
				Family:Cap-Gly domain
12	d1ulia1	6.6	19	Fold:ISP domain
				Superfamily:ISP domain
				Family:Ring hydroxylating alpha subunit ISP domain
13	c2i7fB_	6.5	33	PDB header:oxidoreductase
				Chain: B: PDB Molecule:ferredoxin component of dioxygenase;
				PDBTitle: sphingomonas yanoikuyae b1 ferredoxin
14	d2e3ha1	6	53	Fold:SH3-like barrel
				Superfamily:Cap-Gly domain
				Family:Cap-Gly domain
15	d1ks0a_	5.8	63	Fold:Kringle-like
				Superfamily:Kringle-like
				Family:Fibronectin type II module
16	d1ixda_	5.7	27	Fold:SH3-like barrel
				Superfamily:Cap-Gly domain
				Family:Cap-Gly domain
17	d2gxfa1	5.6	43	Fold:Cystatin-like
				Superfamily:NTF2-like
				Family:YybH-like
18	c3dqyA_	5.4	44	PDB header:oxidoreductase
				Chain: A: PDB Molecule:toluene 1,2-dioxygenase system ferredoxin
				PDBTitle: crystal structure of toluene 2,3-dioxygenase ferredoxin
19	d1knwa1	5.3	24	Fold:Domain of alpha and beta subunits of F1 ATP synthase-like
				Superfamily:Alanine racemase C-terminal domain-like
				Family:Eukaryotic ODC-like

Table 2 – Phyre2 results for PLN SEA autocleavage peptide modeling.
Phyre2 returned a list of proteins that could be used as a template for modeling the PLN SEA autocleavage peptide. Each protein is listed by its PDB id, given a confidence score, the percent of matching protein sequence is listed, and some descriptive information.

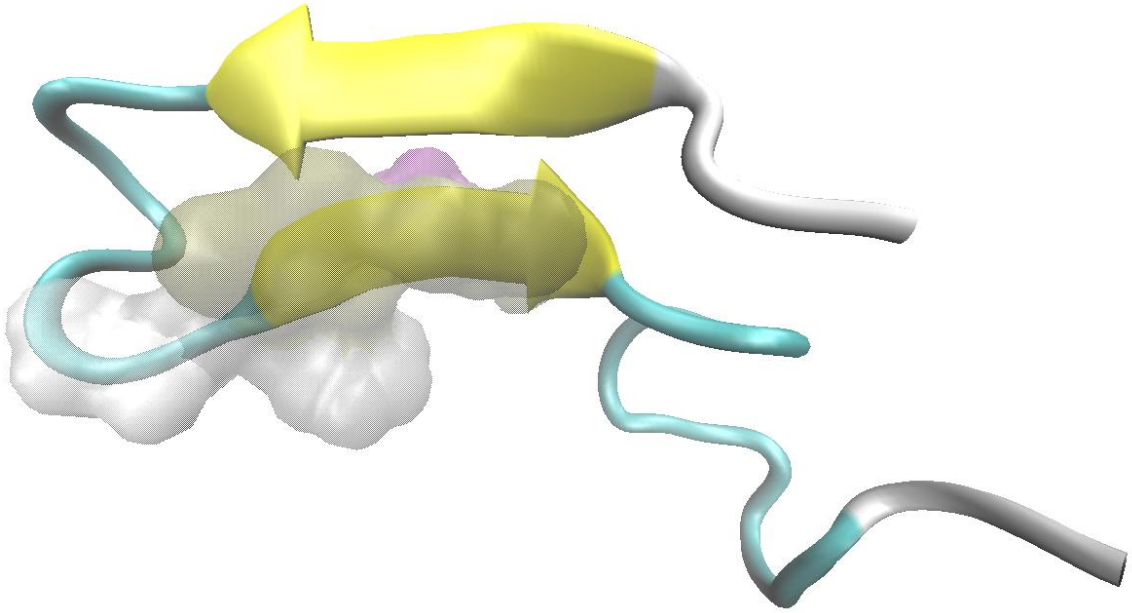
Muc1_SEA_GSVVV Modeling Templates

#	Template	Confidence	% i.d.	Template Information
1	c2acmB	97.2	100	PDB header:structural protein
				Chain: B: PDB Molecule:mucin-1;
				PDBTitle: solution structure of the sea domain of human mucin 1 (muc1)
2	c2acmA	95.8	100	PDB header:structural protein
				Chain: A: PDB Molecule:mucin-1;
				PDBTitle: solution structure of the sea domain of human mucin 1 (muc1)
3	c2e7vA	88.4	29	PDB header:hydrolase
				Chain: A: PDB Molecule:transmembrane protease;
				PDBTitle: crystal structure of sea domain of transmembrane protease2 from mus musculus
4	c1cf1B	50.4	67	PDB header:structural protein
				Chain: B: PDB Molecule:protein (arrestin);
				PDBTitle: arrestin from bovine rod outer segments
5	c1ayrA	40.5	67	PDB header:sensory transduction
				Chain: A: PDB Molecule:arrestin;
				PDBTitle: arrestin from bovine rod outer segments
6	d1g4ma1	30.5	67	Fold:Immunoglobulin-like beta-sandwich
				Superfamily:E set domains
				Family:Arrestin/Vps26-like
7	d1cf1a1	30.4	67	Fold:Immunoglobulin-like beta-sandwich
				Superfamily:E set domains
				Family:Arrestin/Vps26-like
8	c1jsyA_	27.4	67	PDB header:signaling protein
				Chain: A: PDB Molecule:bovine arrestin-2 (full length);
				PDBTitle: crystal structure of bovine arrestin-2
9	c2uvaI_	12.3	62	PDB header:transferase
				Chain: I: PDB Molecule:fatty acid synthase beta subunits;
				PDBTitle: crystal structure of fatty acid synthase from thermomyces2 lanuginosus at 3.1 angstrom resolution. this file contains3 the beta subunits of the fatty acid synthase. the entire4 crystal structure consists of one heterododecameric fatty5 acid synthase and is described in remark 400
10	c2vkzH_	11.3	46	PDB header:transferase
				Chain: H: PDB Molecule:fatty acid synthase subunit beta;
				PDBTitle: structure of the cerulenin-inhibited fungal fatty acid2 synthase type i multienzyme complex

11	c1sujA_	9.5	67	PDB header:signaling protein
				Chain: A: PDB Molecule:cone arrestin;
				PDBTitle: x-ray crystal structure of ambystoma tigrinum cone arrestin
12	d1d2oa1	7.6	50	Fold:Prealbumin-like
				Superfamily:Cna protein B-type domain
				Family:Cna protein B-type domain
13	d1s1da_	7.2	45	Fold:5-bladed beta-propeller
				Superfamily:Apyrase
				Family:Apyrase
14	d1d2oa2	7.1	40	Fold:Prealbumin-like
				Superfamily:Cna protein B-type domain
				Family:Cna protein B-type domain
15	d1ik9a1	5.8	36	Fold:XRCC4, N-terminal domain
				Superfamily:XRCC4, N-terminal domain
				Family:XRCC4, N-terminal domain
16	c1g4mA_	5.2	67	PDB header:signaling protein
				Chain: A: PDB Molecule:beta-arrestin1;
				PDBTitle: crystal structure of bovine beta-arrestin 1

Table 3 - Phyre2 results for MUC1 SEA autocleavage peptide modeling.
Phyre2 returned a list of proteins that could be used as a template for modeling the PLN SEA autocleavage peptide. Each protein is listed by its PDB id, given a confidence score, the percent of matching protein sequence is listed, and some descriptive information.

A PLN SEA



B MUC1 SEA

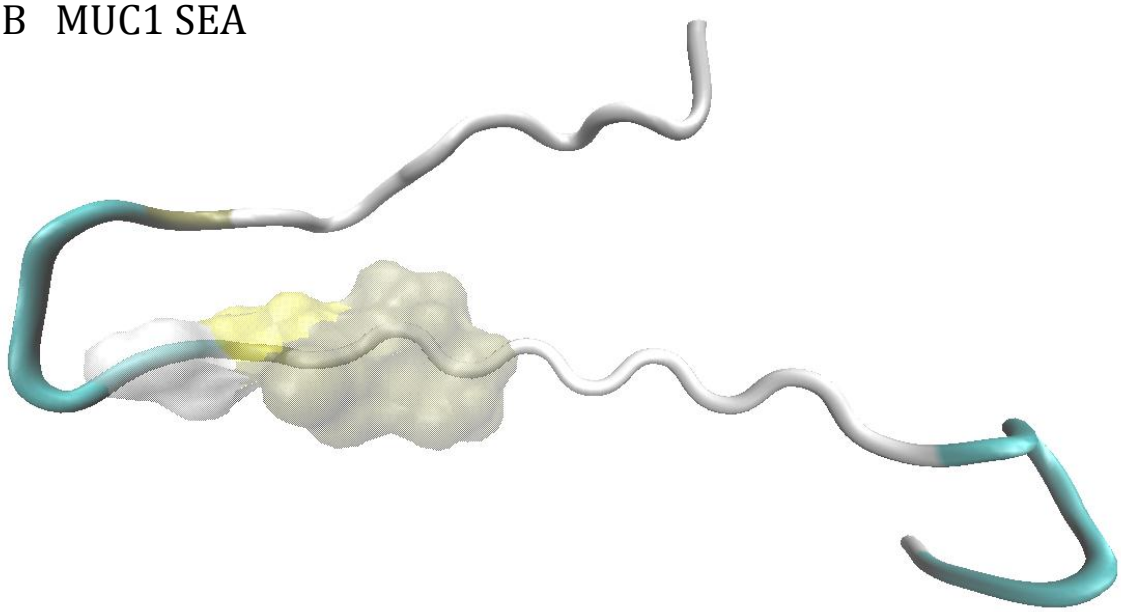


Figure 11 – Model of autocleavage regions of PLN and MUC1 SEAs.
Models created by Phyre2. Structures are displayed in the NewCartoon style using VMD visualization software. Autocleavage motif residues are highlighted as semitransparent space filling residues. A) PLN SEA autocleavage peptide. Residues 62-67 (GWVFV) of the PLN SEA are highlighted, corresponding to residues 142 – 146 of PLN. B) MUC1 SEA autocleavage peptide. Residues 64 – 68 (GSVVV) are highlighted. These correspond to residues 1097 – 1011 of MUC1.

refinement iterations. The information from this alignment combined with information from other SEA modules studies led to a second alignment with SEAs from fewer proteins but more species. The SEA modules from ENTK, GP110, GP116, MUC1, MUC3A, MUC3B, MUC12, MUC17, PLN, UROL1, and the second SEA module of IMPG1 and IMPG2 of different species were analyzed using the same Geneious alignment with Blosum62 score matrix. A third alignment using the same methods above was performed on DAG1, GP116, MUC1, Muc3A, MUC3B, MUC12, and TMPRSS6 for autocleavage predictions.

2.6.2. Folding predictions

The PLN SEA module sequence from table six was used for the folding predictions. No changes or mutations were added to the sequence. The sequence was put into the web based protein folding prediction website Phyre2 (Kelley and Sternberg, 2009). The job was given a unique identifier (1a9df393dfac51a5) and the results were emailed to me. Multiple proteins were returned as possible templates for folding but the one with the highest confidence was used to create the folding model regardless of the percent identity match. The downloaded model was viewed and manipulated using VMD software v1.9.1 (Humphrey et *al.*, 1996). The PLN SEA model and the MUC1 SEA (PDB ID - 2ACM) were both displayed in this software. The molecules were drawn in the New Cartoon style and colored based on the secondary structure. Residues in the autocleavage motif were located in the

Sequence viewer window under the Extensions tab in the Analysis subfolder. The autocleavage residues were visually highlighted as beads.

The same methods were used to evaluate the fold of the short 31 amino acid sequences of the autocleavage region and the models need for autocleavage predictions. The sequences were separately pasted into the Phyre2 program. The resulting models were rendered in VMD. Each model was drawn in the New Cartoon style and colored based on the secondary structure. The autocleavage motif was highlighted through the same procedure, but visually emphasized with a semitransparent surface representation around the residues. This was done to identify the region of potential cleavage while still allowing the fold around the residues to be seen. The prediction models highlight the phenylalanine, glycine, and serine mentioned in section 2.4. This was done to show the proximity of the phenylalanine to the autocleavage region.

2.6.3. Recombinant protein production and isolation

Initially the DNA encoding the PLN and MUC1 SEA modules needed to be isolated and amplified in order to begin the molecular cloning necessary for protein expression. The transcript sequence for PLN (PLN - ENST00000374695) and thus the PLN SEA was obtained from Ensembl (http://useast.ensembl.org/index.html). Using the transcript information custom forward and reverse primers were

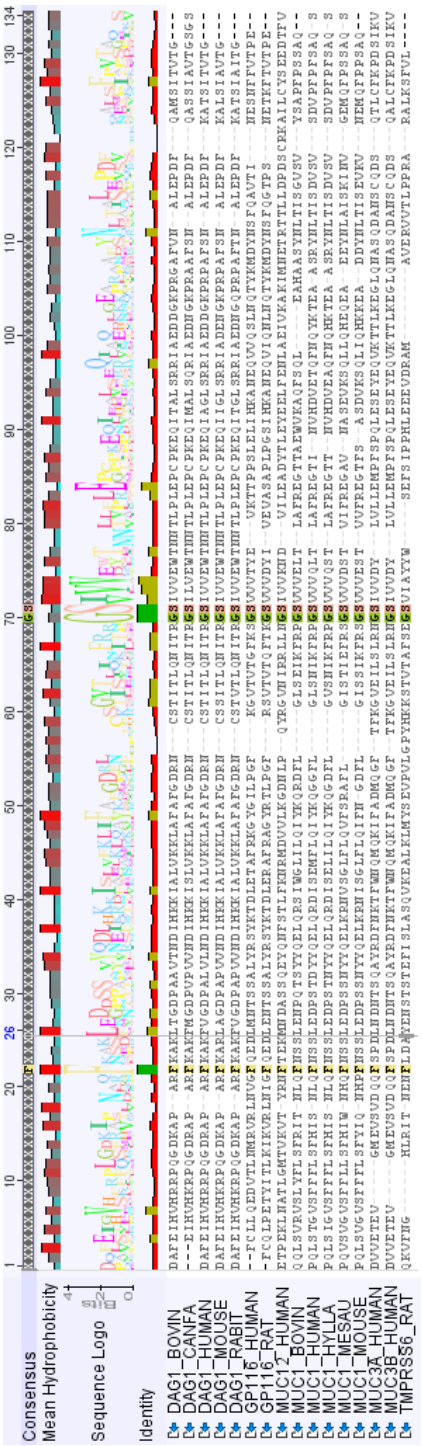


Figure 12 – Sequence alignment of autocleaving SEA modules.
Geneious alignment with a Blosum62 cost matrix and an exception of 1. Highlighted residues have 100% consensus across the sequence alignment. SEAs in this alignment include DAG1, GP116, MUC1, MUC3A, MUC3B, MUC12, and TMPRSS6 from various species.

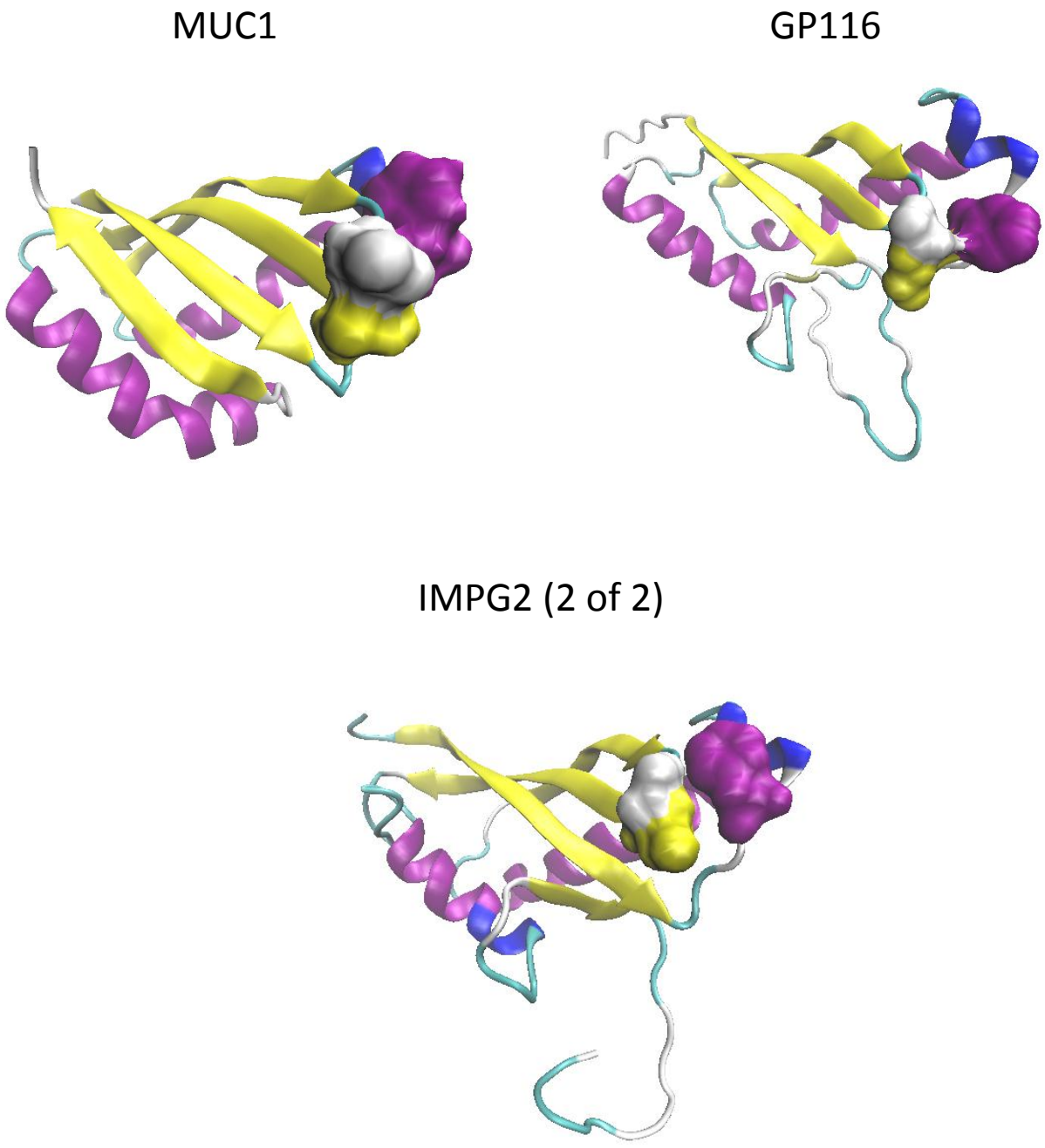


Figure 13 – Modeling location of fully conserved residues in autocleaving SEA modules.
Models created by Phyre2. Structures are displayed in the NewCartoon style using VMD visualization software. Residues noted from figure 12 are highlighted as surface representations. The glycine is colored white. The serine is yellow and the phenylalanine is purple. MUC1 and GP116 SEA modules are known to autocleave. IMPG2 (2 of 2) is predicted to autocleave.

Autocleaves	Does not autocleave	Predicted to autocleave	Predicted not to autocleave
DAG1_Bovine	PLN_Human	GP110_Human	AGRIN_Chicken
DAG1_Dog	PLN_Mouse	GP110_Mouse	AGRIN_Electric ray
DAG1_Human		IMPG1_Bovine(2 of 2)	AGRIN_Human
DAG1_Mouse		IMPG1_Human(2 of 2)	AGRIN_Mouse
DAG1_Rabbit		IMPG1_Mouse(2 of 2)	AGRIN_Rat
GP116_Human		IMPG1_Rat(2 of 2)	DESC4_Mouse
GP116_Rat		IMPG2_Chicken(2 of 2)	DESC4_Rat
MUC01_Bovine		IMPG2_Human(2 of 2)	ENTK_Bovine
MUC01_Gibbon		IMPG2_Mouse(2 of 2)	ENTK_Human
MUC01_Hamster		IMPG2_Rat(2 of 2)	ENTK_Mouse
MUC01_Human		MUC17_Human	ENTK_Pig
MUC01_Mouse		TMPRSS6_Rat	IMPG1_Bovine(1 of 2)
MUC03A_Human			IMPG1_Chicken
MUC03B_Human			IMPG1_Human(1 of 2)
MUC12_Human			IMPG1_Mouse(1 of 2)
			IMPG1_Rat(1 of 2)
			IMPG2_Chicken(1 of 2)
			IMPG2_Human(1 of 2)
			IMPG2_Mouse(1 of 2)
			IMPG2_Rat(1 of 2)
			MUA3_Caeel
			MUC13_Human
			MUC13_Mouse
			MUC13_Rat
			SP63_Sea urchin
			TM11A_Human
			TM11A_Mouse
			TM11B_Human
			TM11B_Mouse
			TM11B_Rat
			TM11D_Human
			TM11D_Mouse
			TM11D_Rat
			TM11E_Human
			TM11E_Mouse
			TM11F_Human
			TM11F_Mouse
			UROL1_Human(1of2)
			UROL1_Human(2of2)

Table 4 – Autocleavage predictions.
The SEAs list as ones the autocleave were identified from literature. PLN SEA modules does not autocleave based on this work. The SEAs predicted to autocleave have both the phenylalanine and the glycine/serine of the autocleavage motif identified in figure12 and highlighted in figure 13. The SEAs predicted not to autocleave lack one or both of the features.

designed and ordered from Integrated DNA Technologies to amplify each of the SEA sequences.

Two sets of primers were designed for each SEA module. The set used to create an N-terminally tagged expression construct removed the start codon from the initial SEA sequence to later add it before the HIS tag by the recombination event with the destination vector. This set of primers also maintained the stop codon from the original sequence. The set of primers used to create the C-terminally tagged expression construct removed the stop codon from the sequence so that it could be added back after the HIS tag by the destination vector. This set of primers also maintained the start codon from the original sequence.

cDNA from a pancreatic epithelial cell line (HPAF), kindly provided by another graduate student in the laboratory, Patricia Chapela, was used as the template for the initial PCR reaction. The PCR reaction was combined and cycled using a typical PCR protocol. The products were separated on an 1.5% (w/v) agarose gel containing 1µl of 10 mg/mL ethidium bromide. Bands that corresponded to the correct sized were cut from the gel and the DNA was extracted with a Qiagen gel extraction kit. All DNA was sequenced by LoneStar labs (Houston, TX) before moving forward with the cloning process.

The sequence verified DNA was combined with either the pENTR/SD/D-TOPO or the pENTR/D-TOPO (Life Technologies) depending on whether the goal was to create an N- or C- terminally tagged final product. The reaction was

maintained at 30 minutes at room temperature before being placed on ice. The reaction products were transformed into chemically competent One Shot® TOP10 *E. coli* and plated on kanamycin containing LB agar plates. Positive colonies were selected and grown over night in selective LB broth. The plasmids were isolated using the Qiagen miniprep kit and sequenced by Lonestar labs (Houston, TX). Sequenced verified entry clones and destination clones were combined with the LR Clonase™ II enzyme mix in TE buffer to perform the recombination reaction. The reaction was held at room temperature for one hour before proteinase K was added and the reaction was moved to 37°C for ten minutes. A small volume was transformed into competent cells as before and plated on ampicillin containing LB agar plates. Positive colonies were grown in broth overnight and their plasmid isolated and sequenced as described above. All sequence analysis data from Lonestar Labs (Houston, TX) was analyzed using Geneious software v5.3.6. More information on the vectors used for cloning can be found in figure 15.

To produce protein, destination vectors were transformed into chemically competent *E. coli* designed for the expression of recombinant protein, One Shot® BL21 Star™ (DE3) Chemically Competent *E. coli* (Life Technologies). Transformed cells were inoculated into 5ml 2YT media containing ampicillin and grown overnight in a 37^0C incubator. The overnight broth was inoculated into 1 liter of selective 2xYT media and grown in a 37^0C shaker until OD_{600} = 1.0. 1mM isopropyl

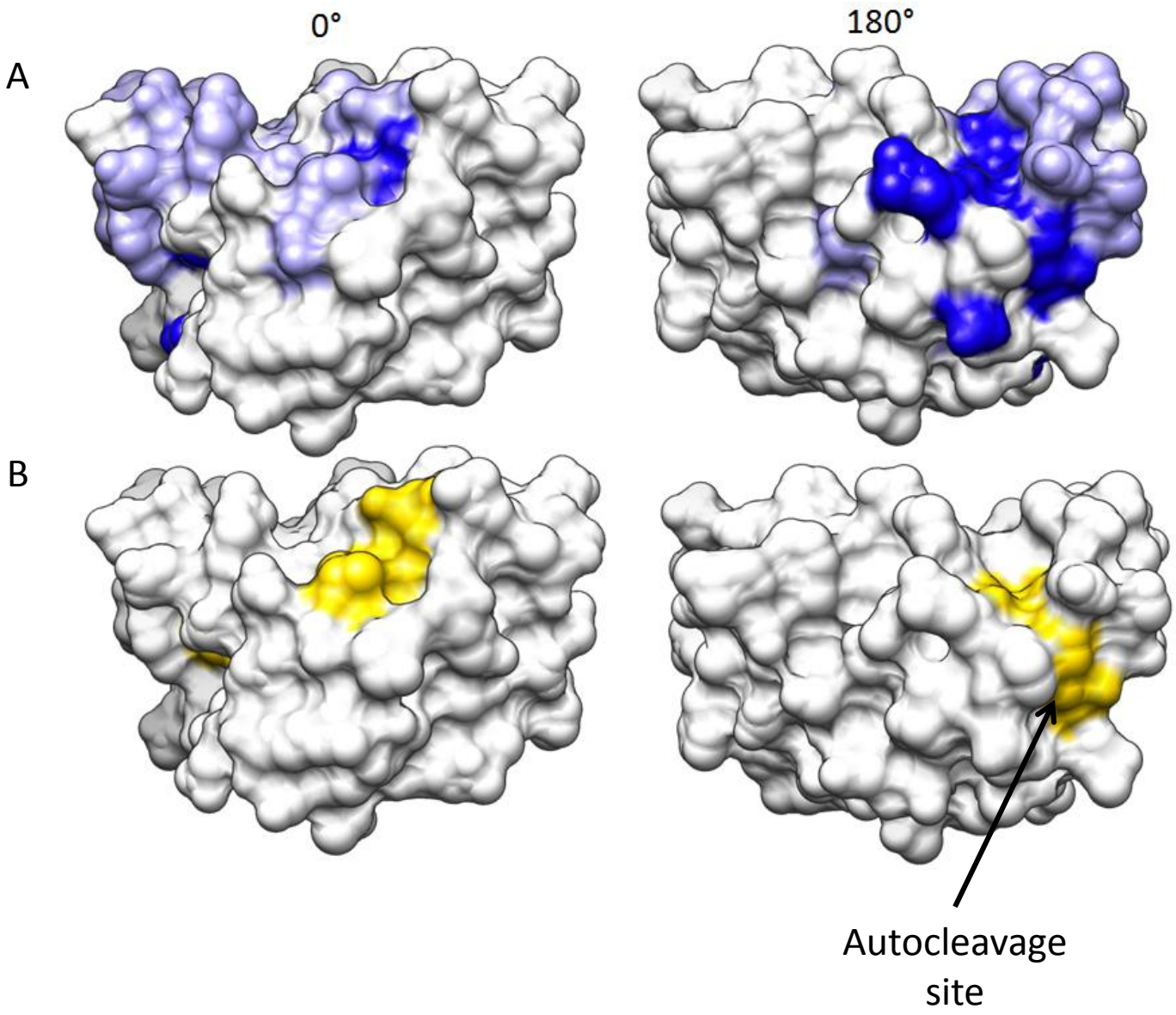


Figure 14 – Space filling SEA model of conserved residues.
A) Blue coloring highlights conserved residues among various SEA modules. Dark blue denotes surface exposed residues. Light blue denotes conserver residues buried in the fold. B) Yellow denoted conserved hydrophobic solvent facing residues. Specific residues are noted by an (star) on figure 6. Images made in collaboration with Drew Bryant (see text).

β-D-1-thiogalactopyranoside (IPTG) (0.238g) was added to the culture to induce protein production. The culture was left in the incubator overnight. To collect the protein produced the culture was first spun down at 10,000 rpm for 30 minutes and the supernatant removed. The cell pellet was lysed through a series of freeze thaw cycles in liquid nitrogen. After each cycle 10ml of buffer supplemented with 5mM phenylmethylsulfonyl fluoride, 5mM N_{α}-*p*-Tosyl-L-arginine methyl ester hydrochloride, and 100µg/ml lysozyme. After the third freeze/thaw 10ml of buffer containing 0.01g of DNase was added and the solution was allowed to sit on ice until it was fluid. The lysate was then centrifuged at 12,000 rpm for 30 minutes and the supernatant or cell lysate was collected. The whole cell lysate was stored at 4°C.

Recombinant protein was isolated from the whole cell lysate using metal ion chromatography. The HIS tagged SEA modules were bound to the Ni-NTA column by gently tumbling the lysate in the column for one hour at 4°C. The rest of the lysate was applied to the column. The column was then washed with a series of low concentration imidazole buffers (50mM NaH_2PO_4, 300mM NaCl, 10-20mM imidazole, pH – 8). The protein of interest was eluted with a high concentration (300mM) imidazole buffer. Protein samples were stored 4°C with 0.02% (w/v) sodium azide. Precipitated protein was subjected to ten 30 second rounds of bore sonication while on ice in order to resolubilize the material in the extract.

Purified recombinant protein stability in different pHs was tested prior to further analysis. Proteins were dialyzed in a 1kDa molecular weight cut off Tube-O-DIALYZER (G Biosciences) into elution buffer of pHs ranging from five to eight. The protein was dialyzed overnight in 4°C. Samples from each pH were electrophoresed on a gradient NuPAGE® 4-12% Bis Tris gel (Life Technologies). The gel was Coomassie stained overnight and destained overnight at room temperature. Alterations in the proteins due to the pH if any were compared sample to sample.

2.6.4. Autocleavage analysis

Autocleavage was analyzed via two different methods, first by coomassie stain SDS-PAGE gels and second by western blot. For the Coomassie analysis two different samples had to be taken from each protein expression procedure. The first sample was removed from the 1 liter culture just before IPTG was added and protein production begins. The second sample was removed after protein production just before the cells were pelleted. The two small samples were centrifuged and the supernatant discarded. The pellet was resuspended in 100µl of 8M urea buffer. The sample was recentrifuged and the cell debris was discarded, saving only the supernatant. These samples were electrophoresed on a gradient SDS-PAGE gel to separate the proteins found in the whole cell lysate. The gel was Coomassie stained overnight at room temperature and then de-stained overnight at room temperature. Gels were imaged on a Kodak Image Station 4000MM PRO from

ID	Species	Accession Number	Length (AA)	Name or Description
AGRIN_HUMAN	*Homo sapiens*	O00468	2045	Agrin
ENTK_BOVIN	*Bos taurus*	P98072	1035	Enteropeptidase
				Enterokinase
				Serine protease 7
				Transmembrane protease serine 15
ENTK_HUMAN	*Homo sapiens*	P98073	1019	Enteropeptidase
				Enterokinase
				Serine protease 7
				Transmembrane protease serine 15
ENTK_MOUSE	*Mus musculus*	P97435	1069	Enterokinase
				Enteropeptidase
				Serine protease 7
				Transmembrane protease serine 15
ENTK_PIG	*Sus scrofa*	P98074	1034	Enteropeptidase
				Enterokinase
				Serine protease 7
				Transmembrane protease serine 15
GP110_HUMAN	*Homo sapiens*	Q5T601	910	Probable G-protein coupled receptor 110
				G-protein coupled receptor KPG_012
				G-protein coupled receptor PGR19
GP110_MOUSE	*Mus musculus*	Q8VEC3	908	G-protein coupled receptor 110
GP116_HUMAN	*Homo sapiens*	Q8IZF2	1346	Probable G-protein coupled receptor 116
GP116_RAT	*Rattus norvegicus*	Q9WVT0	1349	Probable G-protein coupled receptor 116
				G-protein coupled hepta-helical receptor Ig-hepta
IMPG1_BOVIN	*Bos taurus*	Q9GMS5	794	Interphotoreceptor matrix proteoglycan 1
				Mucin-like glycoprotein associated with photoreceptor cells
				Sialoprotein associated with cones and rods
IMPG1_HUMAN	*Homo sapiens*	Q17R60	797	Interphotoreceptor matrix proteoglycan 1
				Interphotoreceptor matrix proteoglycan of 150 kDa
				Sialoprotein associated with cones and rods
IMPG1_MOUSE	*Mus musculus*	Q8R1W8	798	Interphotoreceptor matrix proteoglycan 1
				Interphotoreceptor matrix proteoglycan of 150 kDa
				Sialoprotein associated with cones and rods
IMPG1_RAT	*Rattus norvegicus*	Q9ET62	798	Interphotoreceptor matrix proteoglycan 1
				Mucin-like glycoprotein associated with photoreceptor cells
				Sialoprotein associated with cones and rods

IMPG2_CHICK	*Gallus gallus*	Q1XI86	1423	Interphotoreceptor matrix proteoglycan 2
				Sialoprotein associated with cones and rods proteoglycan
				Sparcan
IMPG2_HUMAN	*Homo sapiens*	Q9BZV3	1241	Interphotoreceptor matrix proteoglycan 2
				Interphotoreceptor matrix proteoglycan of 200 kDa
				Sialoprotein associated with cones and rods proteoglycan or Sparcan
IMPG2_MOUSE	*Mus musculus*	Q80XH2	1243	Interphotoreceptor matrix proteoglycan 2
				Sparcan
IMPG2_RAT	*Rattus norvegicus*	P70628	1241	Interphotoreceptor matrix proteoglycan 2
				Sparcan
MUC1_BOVIN	*Bos taurus*	Q8WML4	580	Mucin-1
				CD_antigen=CD227
MUC1_HUMAN	*Homo sapiens*	P15941	1255	Mucin-1
				Breast carcinoma-associated antigen DF3
				Episialin
				H23AG
				Krebs von den Lungen-6
				PEMT
				Peanut-reactive urinary mucin
				Polymorphic epithelial mucin
				Tumor-associated epithelial membrane antigen
				CD_antigen=CD227
MUC1_HYLLA	*Hylobates lar* (gibbon)	Q29435	475	Mucin-1
				CD_antigen=CD227
MUC1_MESAU	*Mesocricetus auratus* (golden hamster)	Q60528	676	Mucin-1
				CD_antigen=CD227
MUC1_MOUSE	*Mus musculus*	Q02496	630	Mucin-1
				Episialin
				CD_antigen=CD227
MUC3A_HUMAN	*Homo sapiens*	Q02505	2541	Mucin-3A
				Intestinal mucin-3A
MUC3B_HUMAN	*Homo sapiens*	Q9H195	901	Mucin-3B
				Intestinal mucin-3B
MUC12_HUMAN	*Homo sapiens*	Q9UKN1	5478	Mucin-12
MUC13_HUMAN	*Homo sapiens*	Q9H3R2	511	Mucin-13
				Down-regulated in colon cancer 1
MUC17_HUMAN	*Homo sapiens*	Q685J3	4493	Mucin-17
				Small intestinal mucin-3
PGBM_HUMAN	*Homo sapiens*	P98160	4391	Basement membrane-specific heparan sulfate proteoglycan core protein
				Perlecan
PGBM_MOUSE	*Mus musculus*	Q05793	3707	Basement membrane-specific heparan sulfate proteoglycan core protein

TM11A_HUMAN	*Homo sapiens*	Q6ZMR5	421	Transmembrane protease serine 11A
				Airway trypsin-like protease 1
				Epidermal type-II transmembrane serine protease
				Esophageal cancer-susceptibility gene 1 protein
TM11B_HUMAN	*Homo sapiens*	Q86T26	416	Transmembrane protease serine 11B
				Airway trypsin-like protease 5
TM11D_HUMAN	*Homo sapiens*	O60235	418	Transmembrane protease serine 11D
				Airway trypsin-like protease
				Transmembrane protease serine 11D non-catalytic chain
TM11E_HUMAN	*Homo sapiens*	Q9UL52	423	Transmembrane protease serine 11E
				Serine protease DESC1
				Transmembrane protease serine 11E2
TM11F_HUMAN	*Homo sapiens*	Q6ZWK6	438	Transmembrane protease serine 11F
				Airway trypsin-like protease 4
UROL1_HUMAN	*Homo sapiens*	Q5DID0	1381	Uromodulin-like 1
				Olfactorin

Table 5 - Protein list.
A list of the proteins containing an SEA module used in sequence alignments is listed with the species specific to it, the accession number, its length, and alternative names. Information on human PLN is highlighted.

SEA Module ID	Length	Position
Sequence		

AGRIN_HUMAN 123 1130-1252
ATKVFQGVLELEGVEGQELFYTPEMADPKSELFGETARSIESTLDDLFRNSDVKKDFRSVRLRDLGPGKSV
RAIVDVHFDPTTAFRAPDVARALLRQIQVSRRRSLGVRRPLQEHVRFMDFDW

ENTK_BOVIN 116 54-169
KSHEARGTLKIISGATYNPHLQDKLSVDFKVLAFDIQQMIDDIFQSSNLKNEYKNSRVLQFENGSIIVIFD
LLFDQWVSDKNVKEELIQGIEANKSSQLVTFHIDLNSIDITASLE

ENTK_HUMAN 118 52-169
LGQSHEARATFKITSGVTYNPNLQDKLSVDFKVLAFDLQQMIDEIFLSSNLKNEYKNSRVLQFENGSIIVV
FDLFFAQWVSDENVKEELIQGLEANKSSQLVTFHIDLNSVDILDKLT

ENTK_MOUSE 118 52-169
LGKSHEVRGTFKITSGVTYNPNLQDKHSVDFKVLAFDLQQMIDEIFESSSLKNEYEKSKVFQFEKGSVIVL
FDLFFAQWVSDKNVKEELIQGIEANISSQLVTLHIDLNSIDITASLS

ENTK_PIG 118 52-169
LGKSHEARGTMKITSGVTYNPNLQDKLSVDFKVLAFDIQQMIGEIFQSSNLKNEYKNSRVLQFENGSVIVI
FDLLFAQWVSDENIKEELIQGIEANKSSQLVAFHIDVNSIDITESLE

GP110_HUMAN 107 146-252
FCERTKIWGTFKINERFTNDLLNSSSAIYSKYANGIEIQLKKAYERIQGFESVQVTQFRNGSIVAGYEVVG
SSSASELLSAIEHVAEKAKTALHKLFPLEDGSFRVF

GP110_MOUSE 107 145-251
FCERAKVWGTFEIDEKFPEDLWNSSSDVYAHYTVGIENQLKEAYRRVHGFESVRVTQFRKGSIVVGYEVTG
STSPPELLFAIEQEAEKAQEALRRQFPVKYGSFRVF

GP116_HUMAN 119 163-281
FCLLQEDVTLNMRVRLNVGFQEDLMNTSSALYRSYKTDLETAFRKGYGILPGFKGVTVTGFKSGSVVVTYE
VKTTPPSLELIHKANEQVVQSLNQTYKMDYNSFQAVTINESNFFVTPE

GP116_RAT 121 159-279
FCQLPETYITLKIKVRLNIGFQEDLENTSSALYRSYKTDLERAFRAGYRTLPGFRSVTVTQFTKGSVVVDY
IVEVASAPLPGSIHKANEQVIQNLNQTYKMDYNSFQGTPSNETKFTVTPE

IMPG1-2_BOVIN 119 575-693
AARGRELVVFFSLRVANVPFSTDLFNKSSLEYQALEQRFTQLLVPNLRSNLTGFKQLEILNFRNGSVIVNS
KVRFAKSVPYNLTKAVRGVLEDFRSTAAQQLDLEIDSYSLDVEPADQA

IMPG1-1_HUMAN 116 230-345
VLEEQRVELSVSLVNQKFKAELADSQSPYYQELAGKSQLQMQKIFKKLPGFKKIHVLGFRPKKEKDGSSST
EMQLTAIFKRHSAEAKSPASDLLSFDSNKIESEEVYHGTMEEDKQ

IMPG1-2_HUMAN 119 567-685
APKGRELVVFFSLRVANMAFSNDLFNKSSLEYRALEQQFTQLLVPYLRSNLTGFKQLEILNFRNGSVIVNS
KMKFAKSVPYNLTKAVHGVLEDFRSAAAQQLHLEIDSYSLNIEPADQA

IMPG1-2_MOUSE 119 571-689
ATKGQELVVFFSLRVANMPFSYDLFNKSSLEYQALEQRFTDLLVPYLRSNLTGFKQLEILSFRNGSVIVNS
KVRFAKAVPYNLTQAVRGVLEDLRSTAAQGLNLEIESYSLDIEPADQA

IMPG1-2_RAT 119 570-688
AAKGHELVVFFSLRVANMPFSYDLFNKSSLEYQALEQRFTDLLVPYLRSNLTGFKQLEILSFRNGSVIVNS
KVRFAKAVPYNLTQAVRGVLEDLRSTAAQELNLEIESYSLDIEPADQA

IMPG2-2_CHICK 119 1079-1197
PVPSRALVVFFSLRVTNMMFSEDLFNKNSPEYKALEQRFLELLVPYLQSNLTGFQNLEILNFRNGSIVVNS
RMKFAKPVPRNVTNAVYMILEDFCNTAYHTMNLAIDKYSLDVESGEQA

IMPG2-1_HUMAN 123 234-356
ATKPAGEQIAEFSIHLLGKQYREELQDSSSFHHQHLEEEFISEVENAFTGLPGYKEIRVLEFRSPKENDSG
VDVYYAVTFNGEAISNTTWDLISLHSNKVENHGLVELDDKPTVVYTISNFRD

IMPG2-2_HUMAN 120 892-1011
YTQTSGALVVFFSLRVTNMMFSEDLFNKNSLEYKALEQRFLELLVPYLQSNLTGFQNLEILNFRNGSIVVN
SRMKFANSVPPNVNNAVYMILEDFCTTAYNTMNLAIDKYSLDVESGDEA

IMPG2-2_MOUSE 122 895-1016
HTQTAGALVVFFSLRVTNMLFSEDLFNKNSLEYKALEQRFLELLVPYLQSNLSGFQNLEILSFRNGSIVVN
SRVRFAESAPPNVNKAMYRILEDFCTTAYQTMNLDIDKYSLDVESGDEANP

IMPG2-2_RAT 124 895-1018
RTQTAGALVVFFSLRVTNMLFSEDLFNKNSLEYKALEQRFLELLVPYLQSNLSGFQNLEILNFRNGSIVVN
SRVKFAESVPPNVNNAIYMILEDFCTTAYQTMNLDIDKYSLDVESGDDANPCK

MUC1_BOVIN 116 363-478
QQLSVRVSLYFLSFRITNLQFNSSLENPQTSYYQELQRSIWGLILQIYKQRDFLGLSEIKFRPGSVVVELT
LAFREGTTAEWVKAQFSQLEAHAASYNLTISGVSVYSAPFPSSAQ

MUC1_HUMAN 119 1034-1152
PQLSTGVSFFFLSFHISNLQFNSSLEDPSTDYYQELQRDISEMFLQIYKQGGFLGLSNIKFRPGSVVVQLT
LAFREGTINVHDVETQFNQYKTEAASRYNLTISDVSVSDVPFPFSAQS

MUC1_HYLLA 119 254-372
PQLSIGVSFFFLSFHISNLQFNSSLEDPSTNYYQELQRDISELILQIYKQGDFLGVSNIKFRPGSVVVQST
LAFREGTTNVHDVEAQFNQHKTEAASRYNLTISDVSVSDVPFPFSAQS

MUC1_MESAU 117 458-574
PQVSVGVSFFLLSFHIWNHQFNSSLEDPSSNYYQELKRNVSGLFLQVFSRAFLGISTIEFRSGSVVVDSTV
IFREGAVNASEVKSQLLQHEQEAEEYNLAISKINVGEMQFPSSAQS

MUC1_MOUSE 116 411-526
PQLSVGVSFFFLSFYIQNHPFNSSLEDPSSNYYQELKRNISGLFLQIFNGDFLGISSIKFRSGSVVVESTV
VFREGTFSASDVKSQLIQHKKEADDYNLTISEVKVNEMQFPPSAQ

MUC12_HUMAN 128 5163-5290
ETPEKLNATLGMTVKVTYRNFTEKMNDASSQEYQNFSTLFKNRMDVVLKGDNLPQYRGVNIRRLLNGSIVV
KNDVILEADYTLEYEELFENLAEIVKAKIMNETRTTLLDPDSCRKAILCYSEEDTFV

MUC13_HUMAN 110 211-320
KGKVFPGKISVTVSETFDPEEKHSMAYQDLHSEITSLFKDVFGTSVYGQTVILTVSTSLSPRSEMRADDKF
VNVTIVTILAETTSDNEKTVTEKINKAIRSSSSNFLNYD

MUC17_HUMAN 115 4182-4296
ETISAQMELTVTVTSVKFTEELKNHSSQEFQEFKQTFTEQMNIVYSGIPEYVGVNITKLRLGSVVVEHDVL
LRTKYTPEYKTVLDNATEVVKEKITKVTTQQIMINDICSDMMCF

MUC3A_HUMAN 121 2236-2356
DVVETEVGMEVSVDQQFSPDLNDNTSQAYRDFNKTFWNQMQKIFADMQGFTFKGVEILSLRNGSIVVDYLV
LLEMPFSPQLESEYEQVKTTLKEGLQNASQDANSCQDSQTLCFKPDSIKV

MUC3B_HUMAN 121 596-716
DVVETEVGMEVSVDQQFSPDLNDNTSQAYRDFNKTFWNQMQKIFADMQGFTFKGVEILSLRNGSIVVDYLV
LLEMPFSPQLESEYEQVKTTLKEGLQNASQDANSCQDSQALCFKPDSIKV

PLN_HUMAN 115 80-194
QMVYFRALVNFTRSIEYSPQLEDAGSREFREVSEAVVDTLESEYLKIPGDQVVSVVFIKELDGWVFVELDV
GSEGNADGAQIQEMLLRVISSGSVASYVTSPQGFQFRRLGTVPQ

PLN_MOUSE 115 80-194
QMVYFRALVNFTRSIEYSPQLEDASAKEFREVSEAVVEKLEPEYRKIPGDQIVSVVFIKELDGWVFVELDV
GSEGNADGSQIQEVLHTVVSSGSIGPYVTSPWGFKFRRLGTVPQ

TM11A_HUMAN 123 45-167
DQKKEYYHGSFKILDPQINNNFGQSNTYQLKDLRETTENLVSQVDEIFIDSAWKKNYIKNQVVRLTPEEDG
VKVDVIMVFQFPSTEQRAVREKKIQSILNQKIRNLRALPINASSVQVNAMSS

TM11B_HUMAN 123 38-160
FLAVEKTYYYQGDFHISGVTYNDNCENAASQASTNLSKDIETKMLNAFQNSSIYKEYVKSEVIKLLPNANG
SNVQLQLKFKFPPAEGVSMRTKIKAKLHQMLKNNMASWNAVPASIKLMEISK

TM11D_HUMAN 121 44-164
DQKSYFYRSSFQLLNVEYNSQLNSPATQEYRTLSGRIESLITKTFKESNLRNQFIRAHVAKLRQDGSGVRA
DVVMKFQFTRNNNGASMKSRIESVLRQMLNNSGNLEINPSTEITSLTDQA

TM11E_HUMAN 122 44-165
RYNQKKTYNYYSTLSFTTDKLYAEFGREASNNFTEMSQRLESMVKNAFYKSPLREEFVKSQVIKFSQQKHG
VLAHMLLICRFHSTEDPETVDKIVQLVLHEKLQDAVGPPKVDPHSVKIKKI

TM11F_HUMAN 127 52-178
VVEDDKSFYYLASFKVTNIKYKENYGIRSSREFIERSHQIERMMSRIFRHSSVGGRFIKSHVIKLSPDEQG
VDILIVLIFRYPSTDSAEQIKKKIEKALYQSLKTKQLSLTINKPSFRLTPIDSKKM

UROL1-1_HUMAN 112 395-506
TNAQVFEVTIKIVNHNLTEKLLNRSSVEYQDFSRQLLHEVESSFPPVVSDLYRSGKLRMQIVSLQAGSVVV
RLKLTVQDPGFPMGISTLAPILQPLLASTVFQIDRQGTRVQ

UROL1-2_HUMAN 104 787-890
TAARKLIGKVRIKNVRYSESFRNASSQEYRDFLELFFRMVRGSLPATMCQHMDAGGVRME
VVSVTNGSIVVEFHLLIIADVDVQEVSAAFLTAFQTVPLLEVIR

Table 6 - SEA list.
A list of the SEA modules used in the sequence alignments. The length of the module is given with its position in the parent protein. The sequence used in the alignments is stated with each SEA module.

Carestream HEALTH (Rochester, New York) and large bands found in the post induction sample and not the pre induction sample were identified as protein of interest. Their size was compared to the predicted molecular weight of the recombinant protein. If a band electrophoresed faster than expected the SEA would be considered to autocleave and if the protein migrated at the expected size, it can be concluded that no autocleavage occurred.

The second method to analyze autocleavage was western blot analysis. For this method only the purified protein from the nickel column was used and not the whole cell lysate. Final proteins are electrophoresed on gels described above and transferred to nitrocellulose membrane overnight at 40 volts for five hours. The membrane was blocked in 3% (w/v) BSA in PBS with 1% (v/v) Tween buffer for three hours at room temperature and then rinsed with PBS-Tween. The anti-HIS primary mouse antibody recognizes a sequence of four HISs. The membrane was incubated in primary antibody diluted 1:20,000 in blocking buffer for three hours at room temperature and rinsed with PBS-Tween. The membrane was incubated in anti-mouse horseradish peroxidase (HRP) secondary antibody diluted 1:50,000 in blocking buffer for one hour at room temperature. Enhanced chemiluminescence (ECL) reagent was added to the blot in equal parts and incubated for five minutes. Excess ECL reagent was removed from the membrane and the membrane was imaged either on the Kodak Image Station 4000MM PRO or on autoradiography film.

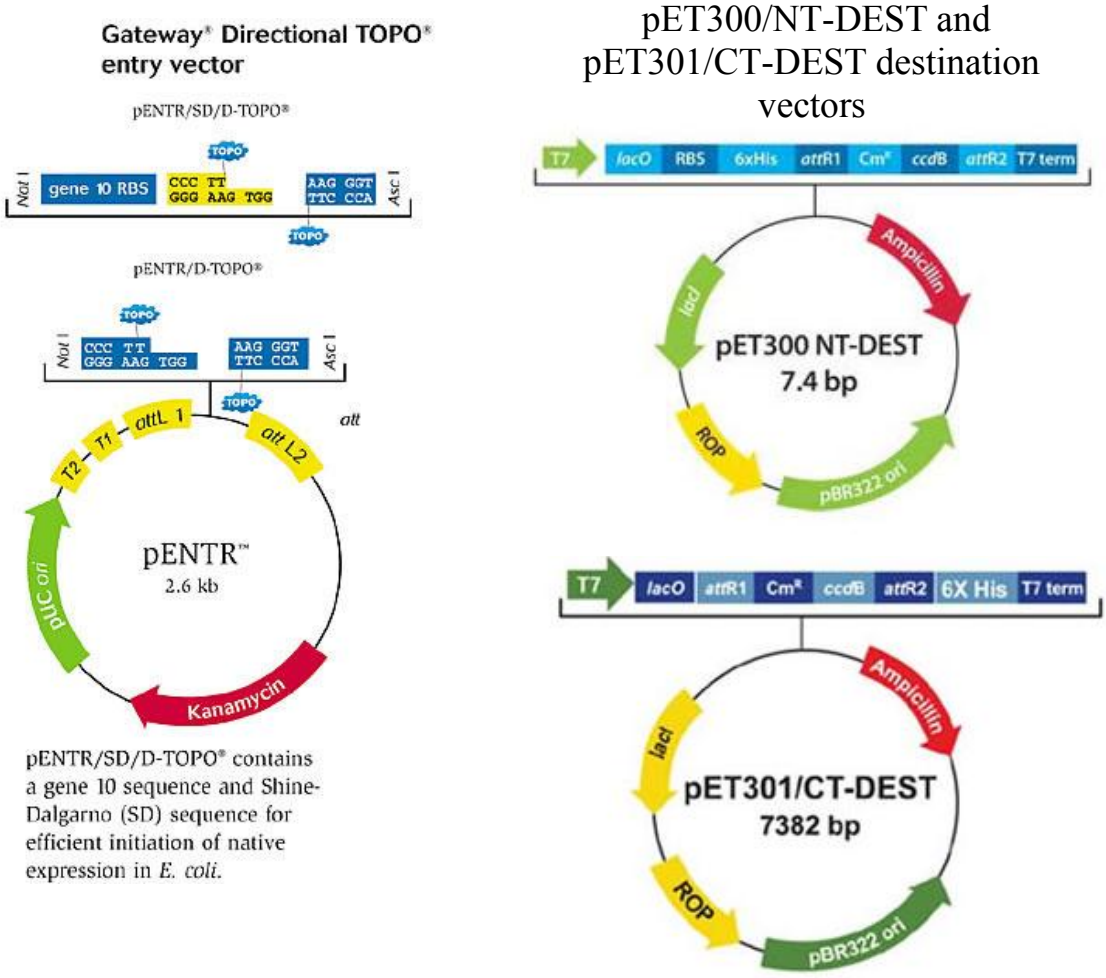


Figure 15 –Vectors.
Information and cartoons of the vectors used in the expression of recombinant PLN and MUC1 SEA modules. These images are courtesy of Life Technologies.
http://products.invitrogen.com/ivgn/product/K242020?ICID=search-product
http://products.invitrogen.com/ivgn/product/K630001?ICID=search-product

2.6.5. Identification of binding residues

The second sequence alignment identified several conserved hydrophobic residues. In collaboration with Drew Bryant, all of the residues were mapped to a space filling model the MUC1 SEA structure (PDB ID - 2ACM). We narrowed the selection to the hydrophobic residues that are surface exposed. All of this modeling work was performed with Chimera interactive visualization and analysis software (Pettersen et *al.*, 2004).

Chapter 3

Identification of Perlecan SEA module binding proteins

The identification of hydrophobic patches on the SEA modules surface begs the question: what binds to those patches and what is the possible effect on the SEA module and the parent protein? To this end a yeast two-hybrid (Y2H) screen was performed to identify PLN SEA binding proteins. Immunocytochemistry (ICC) was performed with mammalian cells in an effort to confirm the Y2H results and rule out and procedural artifacts.

3.1. Identification of binding partners

Hybrigenics (Paris, France) performed the full scale Y2H screen using the PLN SEA that I made as the bait and a cDNA library from human bone marrow endothelial cells (HBMEC)(Schweitzer et *al.*, 1997; Zilberfarb et *al.*, 1997) using an

a/α cell-to-cell mating scheme (Figure 15). HBMECs were used for this screen because PLN in highly expressed in the bone marrow and any proteins that interact with the PLN SEA should be present in these cells. Of 100 million fragment interactions tested, 71 positive clones were identified corresponding to eleven genes and variants (Table 7). The four highest confidence proteins are actin filament associated protein 1 (AFAP1), exosome component 9 (EXOSC9), gamma-aminobutyric acid receptor-associated protein (GABARAP), and F-box protein 28 (FBXO28). Of these proteins, I hypothesize that GABARAP is the most likely to be a *bona fide* PLN SEA module binding protein because of its association with vesicles and ER to Golgi transport.

3.2. Verification of binding partners

To confirm hits from the Y2H screen ICC was performed. This approach was chosen rather than protein complex immunoprecipitation to avoid topological artifacts in the results. WiDr cells from our lab stock were chosen for immunohistochemistry because of the large quantities of PLN they produce (Iozzo, 1984). WiDr cells are an adherent colorectal adenocarcinoma epithelial cell line. WiDr cells were stained with antibodies specific to either PLN or GABARAP or both. PLN localization was identified with A76 antibody and shows PLN intracellularly located in vesicles appearing as speckles (Figure 17). This result is as expected, as it is known that PLN trafficks through the secretory pathway prior to secretion.

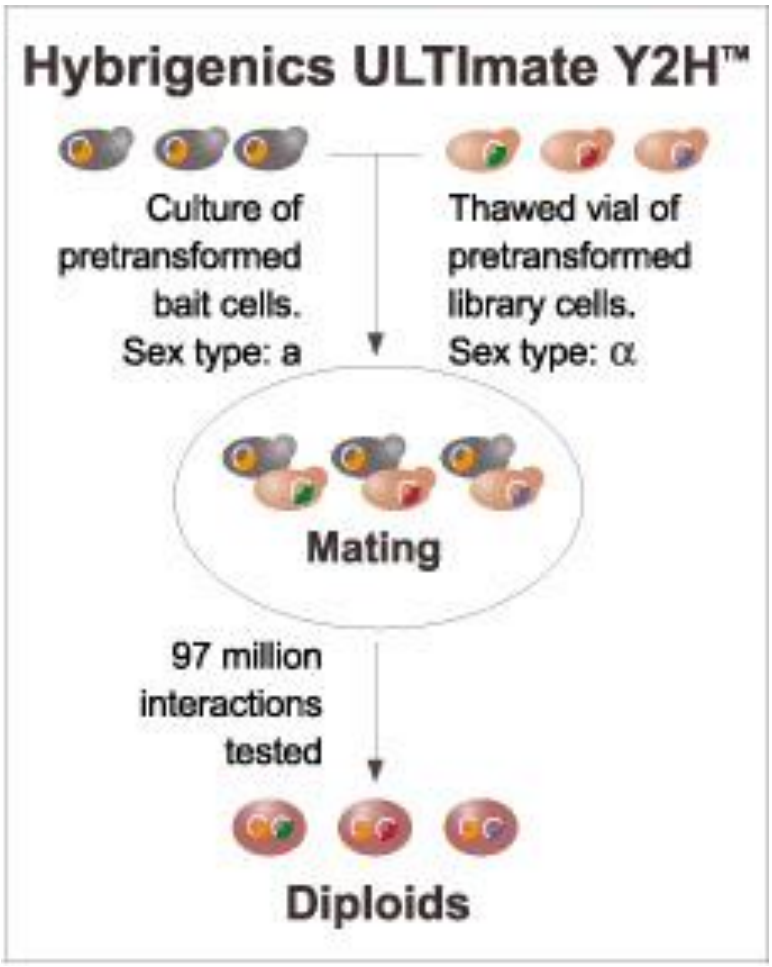


Figure 16 - Yeast two hybrid screen scheme.
This cartoon depicts the general method used by Hybrigenics to run the Y2H screen. Yeast of the a mating type were transformed with the PLN SEA bait plasmid. Yeast of the α mating type were transformed with the HBMEC library of prey plasmids. The two populations were allowed to mate creating diploid cells containing both of the plasmids. Instances where interactions with the bait and prey occurred colonies grew and were evaluated.

PLN
vs Human Bone Marrow Endothelial Cells
Fri, Jul 30, 2010 - 12:47 PM

Screen Parameters
Nature - cDNA
Reference Bait Fragment - HSPG2 (80-194)
Prey Library - Human Bone Marrow Endothelial Cells
Vector - pB27 (N-LexA-bait-C fusion)
Processed Clones - 71 (pB27)
Analyzed Interactions - 111 million

Global PBS (for Interactions represented in the Screen)		#	%
A	Very high confidence in the interaction	3	27.3%
B	High confidence in the interaction	1	9.1%
C	Good confidence in the interaction	0	0.0%
D	Moderate confidence in the interaction This category is the most difficult to interpret because it mixes two classes of interactions: classes of interactions: - False-positive interactions - Interactions hardly detectable by the Y2H technique (low representation of the mRNA in the library, Prey folding, Prey toxicity in yeast)	5	45.5%
E	Interactions involving highly connected prey domains, warning of non-specific interaction. The threshold for high connectivity is 10 for screens with Human, Mouse, Drosophila and Arabidopsis and 6 for all other organisms. They can be classified in different categories: '- Prey proteins that are known to be highly connected due to their biological function '- proteins with a Prey interacting domain that contains a known protein interaction motif or a biochemically promiscuous motif	2	18.2%
F	Experimentally proven technical artifacts	0	0.0%
Non Appliable			
N/A	The PBS is a score that is automatically computed through algorithms and cannot be attributed for the following reasons : - All the fragments of the same reference CDS are antisense - The 5p sequence is missing - All the fragments of the same reference CDS are either All OOF1 or All OOF2 - All the fragments of the same reference CDS lie in the 5' or 3' UTR		

Symbols	Means
*	The fragment contains the full length CDS
↙	Fragment is fully in 5' UTR
↘	Fragment is fully in 3' UTR
x	Fragment contains at least one In Frame STOP codon
[NR]	Fragment was found to be non relevant (poor quality, high N density)
IF OOF1 OOF2	With regard to the theoretical frame of each corresponding CDS (GeneBank), fragments are cloned in frame (IF) if they are in the same frame as Gal4AD. In general, polypeptides synthesized from OOF fragments are not considered of biological interest, unless found together with another frame. However, some of the proteins expressed from an OOF fragment can be translated in the correct frame, due to the

	existence of natural frame-shift events during translation in yeast
??	- the clone sequence is antisense
	- The 5p sequence is missing
N	Antisense
Start..Stop	Position of the 5p and 3p prey fragment ends, relative to the position of the start codon

Clone Name	Type Seq	Gene Name	Start	Stop	Frame	Sense	% Id 5p/3p	PBS
pB27_A-8	5p 3p	AFAP1 var2	996	1864	IF	Sense	97.5 / 98.5	A
pB27_A-52	5p 3p	AFAP1 var2	996	1864	IF	Sense	96.1 / 97.5	A
pB27_A-5	5p 3p	AFAP1 var2	996	1864	IF	Sense	97.1 / 95.5	A
pB27_A-9	5p 3p	AFAP1 var2	996	1864	IF	Sense	98.8 / 97.8	A
pB27_A-10	5p 3p	AFAP1 var2	996	1864	IF	Sense	98.3 / 98.4	A
pB27_A-70	5p 3p	AFAP1 var2	996	1864	IF	Sense	96.2 / 92.6	A
pB27_A-29	5p 3p	AFAP1 var2	1149	1886	IF	Sense	96.9 / 98.9	A
pB27_A-43	5p 3p	AFAP1 var2	1161	1783	IF	Sense	99.7 / 98.7	A
pB27_A-40	5p 3p	AFAP1 var2	1161	1783	IF	Sense	99.5 / 98.4	A
pB27_A-58	5p 3p	AFAP1 var2	1161	1783	IF	Sense	99.7 / 99.5	A
pB27_A-37	5p 3p	AFAP1 var2	1161	1783	IF	Sense	99.7 / 99.5	A
pB27_A-69	5p 3p	AFAP1 var2	1233	1801	IF	Sense	100.0 / 100.0	A
pB27_A-4	5p 3p	AFAP1 var2	1323	1809	IF	Sense	100.0 / 99.8	A
pB27_A-67	5p 3p	AFAP1 var2	1323	1809	IF	Sense	100.0 / 100.0	A
pB27_A-35	5p 3p	AFAP1 var2	1416	1883	IF	Sense	100.0 / 99.4	A
pB27_A-49	5p 3p	AFAP1 var2	1416	1883	IF	Sense	100.0 / 100.0	A
pB27_A-20	5p 3p	AFAP1 var2	1443	1813	IF	Sense	99.7 / 99.7	A
pB27_A-73	5p 3p	AFAP1 var2	1443	1813	IF	Sense	99.7 / 98.9	A
pB27_A-48	5p 3p	AFAP1 var2	1443	1813	IF	Sense	99.7 / 99.7	A
pB27_A-38	5p 3p	AFAP1 var2	1458	1859	IF	Sense	99.8 / 99.8	A
pB27_A-13	5p 3p	AFAP1 var2	1458	1859	IF	Sense	99.8 / 99.8	A
pB27_A-63	5p 3p	AFAP1 var2	1464	1793	IF	Sense	100.0 / 100.0	A
pB27_A-53	5p 3p	AFAP1 var2	1503	1862	IF	Sense	100.0 / 100.0	A
pB27_A-74	5p 3p	AFAP1 var2	1503	1862	IF	Sense	100.0 / 100.0	A
pB27_A-17	5p 3p	AFAP1 var2	1521	1793	IF	Sense	100.0 / 100.0	A
pB27_A-21	5p	AFAP1 varA	1674	2125	IF	Sense	100	D
pB27_A-47	5p 3p	AFAP1 varA	1674	2125	IF	Sense	100.0 / 100.0	D
pB27_A-42	5p 3p	CREBZF	519	1225 x	IF	Sense	99.8 / 98.6	E
pB27_A-51	5p 3p	EXOSC9	6	828	IF	Sense	99.7 / 98.8	A
pB27_A-56	5p 3p	EXOSC9	6	828	IF	Sense	99.9 / 99.9	A
pB27_A-39	5p 3p	EXOSC9	21	902	IF	Sense	98.8 / 98.6	A
pB27_A-1	5p 3p	EXOSC9	69	905	IF	Sense	99.5 / 98.6	A
pB27_A-12	5p	EXOSC9	69	No Data	IF	Sense	99.3	A
pB27_A-45	5p 3p	EXOSC9	72	838	IF	Sense	98.7 / 97.8	A
pB27_A-62	5p 3p	FBXO28	-19	1015	IF	Sense	97.8 / 97.1	B
pB27_A-33	5p 3p	FBXO28	23	1052	OOF2	Sense	97.2 / 94.7	B
pB27_A-57	5p 3p	FBXO28	23	1052	OOF2	Sense	99.6 / 96.2	B
pB27_A-27	5p 3p	FBXO28	23	1052	OOF2	Sense	97.8 / 98.6	B
pB27_A-46	5p 3p	GABARAP	-85	577 *x	IF	Sense	100.0 / 99.0	A

pB27_A-34	5p 3p	GABARAP	-85	577 *x	IF	Sense	100.0 / 99.1	A
pB27_A-19	5p 3p	GABARAP	-70	576 *x	IF	Sense	100.0 / 99.7	A
pB27_A-66	5p 3p	GABARAP	-67	469 *x	IF	Sense	100.0 / 98.2	A
pB27_A-14	5p 3p	GABARAP	-67	454 *x	IF	Sense	99.2 / 99.6	A
pB27_A-22	5p 3p	GABARAP	-55	577 *x	IF	Sense	99.8 / 98.6	A
pB27_A-23	5p 3p	GABARAP	-55	577 *x	IF	Sense	99.8 / 98.5	A
pB27_A-60	5p 3p	GABARAP	-46	571 *x	IF	Sense	100.0 / 99.0	A
pB27_A-16	5p 3p	GABARAP	-46	444 *x	IF	Sense	100.0 / 99.2	A
pB27_A-50	5p 3p	GABARAP	-37	578 *x	IF	Sense	99.8 / 100.0	A
pB27_A-15	5p 3p	GABARAP	-31	455 *x	IF	Sense	99.8 / 93.6	A
pB27_A-44	5p 3p	GABARAP	-31	455 *x	IF	Sense	99.8 / 97.6	A
pB27_A-55	5p 3p	GABARAP	-25	467 *x	IF	Sense	99.8 / 99.4	A
pB27_A-11	5p	GABARAP	-25	467 *x	IF	Sense	100	A
pB27_A-36	5p 3p	GABARAP	-22	579 *x	IF	Sense	100.0 / 98.0	A
pB27_A-68	5p 3p	GABARAP	-22	579 *x	IF	Sense	100.0 / 99.8	A
pB27_A-59	5p 3p	GABARAP	-19	367 *x	IF	Sense	100.0 / 98.7	A
pB27_A-61	5p 3p	GABARAP	-7	574 *x	IF	Sense	100.0 / 99.3	A
pB27_A-64	5p 3p	GABARAP	-7	574 *x	IF	Sense	100.0 / 99.8	A
pB27_A-72	5p 3p	GABARAP	-7	561 *x	IF	Sense	100.0 / 97.6	A
pB27_A-54	5p 3p	GABARAPL1	3	486	IF	Sense	100.0 / 100.0	D
pB27_A-25	5p 3p	ODC1	234	813 x	IF	Sense	100.0 / 97.4	D
pB27_A-26	5p 3p	PLK4	1152	2244	IF	Sense	97.7 / 97.1	D
pB27_A-65	5p 3p	PSMB1	405	836 x	IF	Sense	99.3 / 99.3	D
pB27_A-31	3p	ZNF521	No Data	1483	??	Sense	94.1	E
pB27_A-30	3p	ZNF521	No Data	1483	??	Sense	95.9	E
pB27_A-7	5p 3p	ZNF521	120	1484	IF	Sense	94.3 / 97.1	E
pB27_A-2	5p 3p	ZNF521	120	1484	IF	Sense	95.2 / 99.6	E
pB27_A-18	5p 3p	ZNF521	123	983	IF	Sense	98.4 / 97.9	E
pB27_A-32	5p 3p	ZNF521	123	983	IF	Sense	98.6 / 96.7	E
pB27_A-71	5p 3p	ZNF521	141	973	IF	Sense	96.1 / 91.5	E
pB27_A-28	5p 3p	ZNF521	354	997	IF	Sense	100.0 / 99.5	E
pB27_A-41	5p 3p	ZNF521	378	988	IF	Sense	100.0 / 98.4	E

Table 7 – Y2H screen results.
Type sequence indicates if the 5' and/or 3' sequences were available for prey identification. Gene name indicates the best match from GenBank. Start/Stop indicates the position of the 5p and 3p prey fragment ends, relative to the position of the ATG start codon. % Id 5p/3p indicates the percent identity of the prey fragment sequences with the gene reference sequence. PBS is the confidence of the interaction.

GABARAP staining with the antibody ab109364 shows GABARAP diffusely distributed throughout the cytosol, and also localizing to vesicles (Figure 17). Cells staining with both A76 and ab109364 are questionable as to whether co-localization occurs (Figure 18). The individual antibody staining resembles the single target staining but when the two filters are merged there is little yellow signal that would indicate a co-localization of the two proteins. The two proteins localize to similar regions of the cell but better resolution imaging is needed to conclude co-localization and interaction.

3.3. Materials and Methods

3.3.1. Yeast 2 Hybrid screen

A commercial company, Hybrigenics (Paris, France), was enlisted to perform the Y2H screen and data analysis. The PLN SEA module was used as the bait. The prey in this screen was a library from HBMEC cells provided by the company (Schweitzer et *al.*, 1997; Zilberfarb et *al.*, 1997). I provided template DNA in the form of the PLN SEA entry clone. From this they cloned the SEA sequence into their bait expression vector. The bait and prey vectors were separately transfected into either a or α mating type yeast. Yeast mating resulted in diploid yeast containing both the bait and prey plasmids. Instances where protein interaction between the bait and prey are present colonies grew and data for that prey was analyzed and

could indicate proteins that bind to the bait which in this instance is the PLN SEA module. A more in depth method from Hybrigenics is as follows.

"Yeast two-hybrid screening was performed by Hybrigenics Services, S.A.S., Paris, France (http://www.hybrigenics-services.com). The coding sequence for the human PLN (aa 80-194) was PCR-amplified and cloned into pB27 as a C-terminal fusion to LexA (N-LexA-HSPG2 -C). The construct was checked by sequencing the entire insert and used as a bait to screen a random-primed Human Bone Marrow Endothelial Cells cDNA library constructed into pP6. pB27 and pP6 derive from the original pBTM116 and pGADGH plasmids, respectively (Bartel et *al.*, 1993; Vojtek and Hollenberg, 1995). 111 million clones (12-fold the complexity of the library) were screened using a mating approach with Y187 (mata) and L40DGal4 (mata) yeast strains as previously described (Fromont-Racine et *al.*, 1997). 71 His+ colonies were selected on a medium lacking tryptophan, leucine and HIS. The prey fragments of the positive clones were amplified by PCR and sequenced at their 5' and 3' junctions. The resulting sequences were used to identify the corresponding interacting proteins in the GenBank database (NCBI) using a fully automated procedure. A confidence score (PBS, for Predicted Biological Score) was attributed to each interaction as previously described (Formstecher et *al.*, 2005).

The PBS relies on two different levels of analysis. Firstly, a local score takes into account the redundancy and independency of prey fragments, as well as the distribution of reading frames and stop codons in overlapping fragments. Secondly,

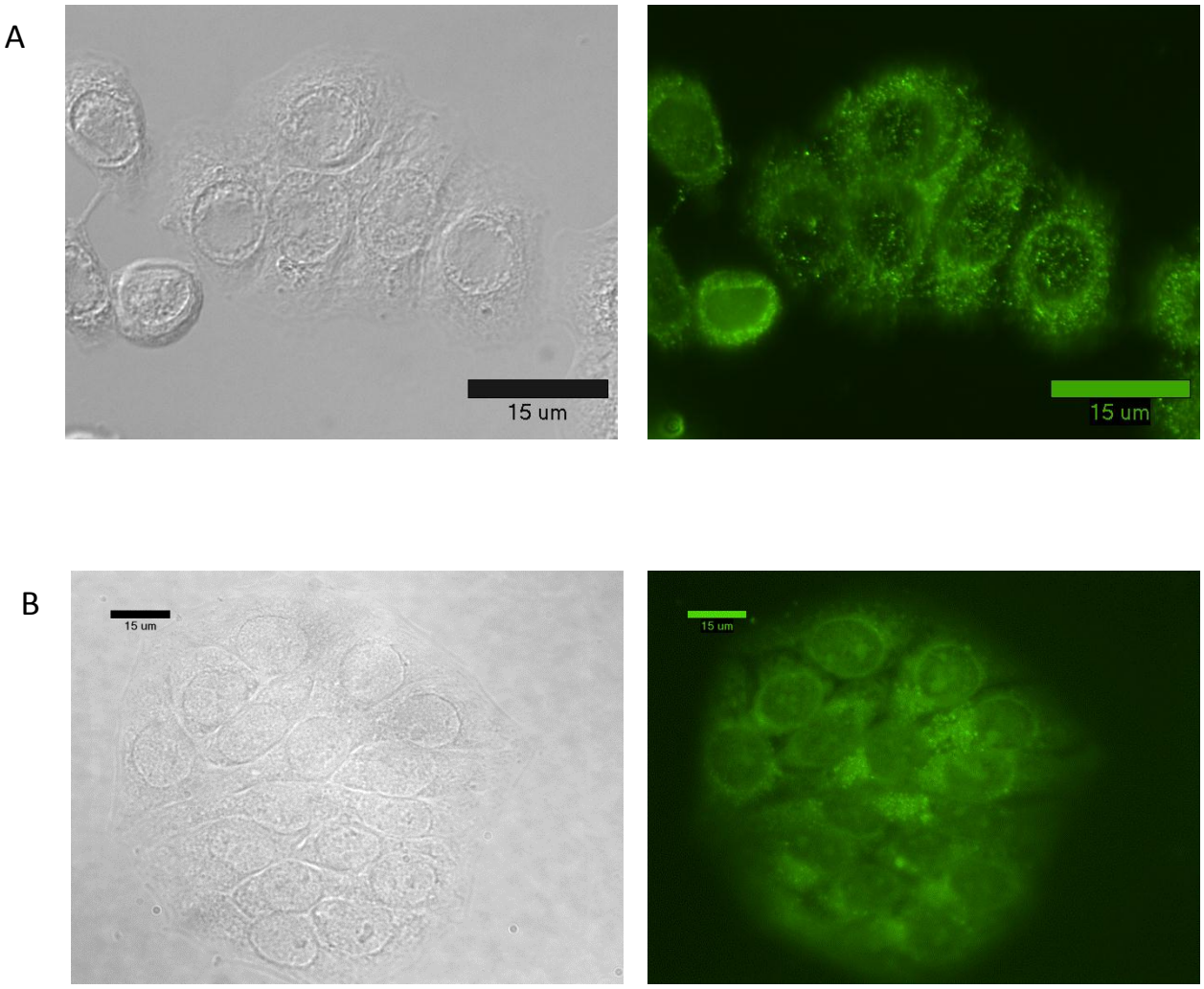


Figure 17 - Single target ICC staining.
WiDr cells were analyzed by indirect immunofluorescence at 100X magnification. Secondary anti-mouse and anti-rabbit Alexa Fluor 488 conjugated antibodies were used. A)PLN was labeled with anti-PLN domain I mouse monoclonal antibody, A76, at a 1:100 concentration. B)GABARAP was labeled with an anti-GABARAP rabbit monoclonal antibody, ab109364, at a concentration of 1:1000.

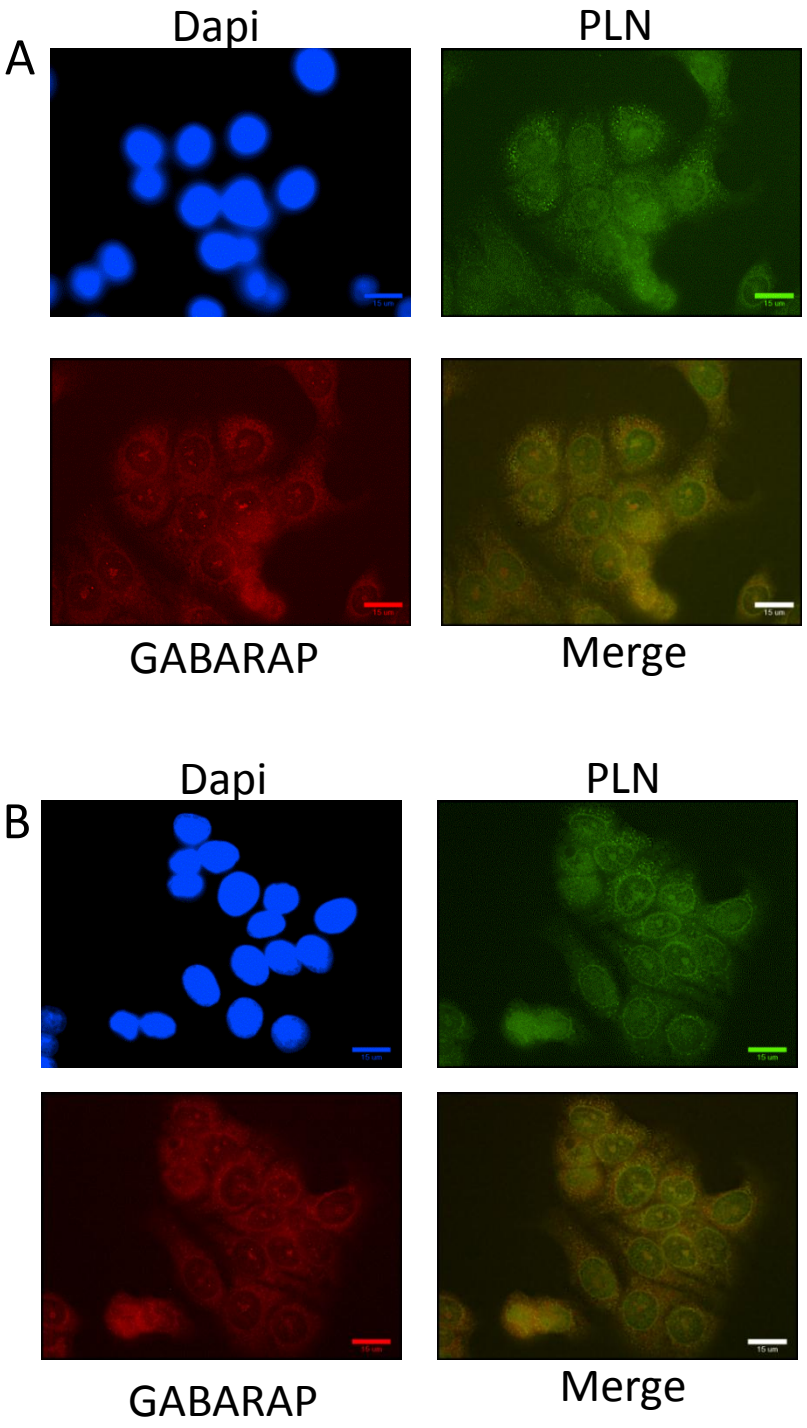


Figure 18 - Co-localization by ICC.
WiDr cells were analyzed by indirect immunofluorescence at 100X magnification. Cells were probed for PLN with anti-PLN domain I mouse monoclonal antibody, A76, at a 1:100 concentration and secondarily labeled with anti-mouse Alexa Fluor 488 conjugated antibodies (green). Cells were probed for GABARAP with and anti-GABARAP rabbit monoclonal antibody, ab109364, at a 1:1000 concentration and secondarily labeled with anti-rabbit Alexa Fluor 647 conjugated antibodies (red). A) and B) are representative images of the staining. Co-localization is difficult to determine. All scale bars are 15µm.

a global score takes into account the interactions found in all the screens performed at Hybrigenics using the same library. This global score represents the probability of an interaction being nonspecific. For practical use, the scores were divided into four categories, from A (highest confidence) to D (lowest confidence). A fifth category (E) specifically flags interactions involving highly connected prey domains previously found several times in screens performed on libraries derived from the same organism. Finally, several of these highly connected domains have been confirmed as false-positives of the technique and are now tagged as F. The PBS scores have been shown to positively correlate with the biological significance of interactions (Rain et *al.*, 2001; Wojcik et *al.*, 2002)."

3.3.2. Immunocytochemistry

WiDr cells were cultured in Dulbecco's Modified Eagle Medium with L-glutamine and supplemented with 10% (v/v) fetal bovine serum (FBS) and 1% (v/v) penicillin streptomycin. Cell were fed twice a week and split at an approximate ratio of 1:6 as recommended by ATCC. For staining cells were grown in 6-well tissue culture plates on standard glass coverslips for the staining procedure. The cells were grown to approximately 60% confluence, when small clusters formed. After that the media was removed and the cells were fixed and permeabilized with ice cold methanol for 30 minutes in -20°C. The cells were briefly rinsed with phosphate buffered saline (PBS) and then blocked in 2.5% (w/v)

BSA in filtered PBS overnight at 4°C. To stain PLN, the primary mouse antibody A76 was used at a concentration of 1:100 (Whitelock et *al.*, 1999). To stain GABARAP the primary rabbit monoclonal ab109364 from Abcam was used at a concentration of 1:1000. Primary antibodies were diluted in 2.5% (w/v) BSA in PBS and incubated overnight in 4°C. The next day the antibody was removed and the cells were washed with PBS. Secondary antibodies were incubated with the cells at a concentration of 1:400 for no less than one hour. Two different secondary antibodies were used in this staining procedure, one tagged with Alexa Flour 488 and another tagged with Alexa Fluor 647. Stained cells were imaged at 100X on an oil objective with either a fluorescein isothiocyanate (Ex: 493nm Em:517nm) or a texas red (Ex: 555nm Em: 615nm) filter on the axioscope microscope in GRB-W using Metamorph software.

Chapter 4

Discussion and Future work

I have shown that overall residue conservation in the SEA modules of various proteins is low and is primarily in the first half of the sequence alignment. The fold of the PLN SEA module is predicted to be similar to that of other SEA modules maintaiing the α/β sandwich. The PLN SEA module does not autocleave and this may be due to sequence features or increased flexibility in the autocleavage region. Proteins that bind to the PLN SEA have been identified and some verification has been performed. In the future more work needs to be done to better characterize the recombinant protein, verify and assess the contribution of binding protein to protein trafficking, and identify which residues are responsible for binding.

4.1. Folding and potential for binding

To examine the folding pattern of the SEA module, the predictive modeling software Phyre2 was used. Multiple possible protein templates were returned with both a confidence score and the percent of matched sequence identity with the PLN SEA module. The template with the highest confidence was used to create the PLN SEA model. Phyre2 predicted the PLN SEA to fold similarly to other SEAs in an α/β sandwich like conformation, often found in RNA binding proteins (Figure 7) (Kranz et *al.*, 1996). The PLN SEA model is similar to that of the MUC1 SEA module. One difference between the two folds is that the PLN SEA has three β strands while the MUC1 SEA has four. The sequence alignments show that most of the conservation is reserved to the first half of the protein sequence. The high similarity, between the two SEA structures, shows that while the sequence conservation from figure 6 was not 100%, enough of the interactions remain to maintain the same fold. This then begs the question: Is the fold enough to cause the PLN SEA module to autocleave in the manner that the MUC1 SEA module does even though the autocleavage sequence is different?

4.2. Comparison of autocleavage in recombinant SEA modules

Autocleavage analysis was performed on recombinantly expressed protein. The analysis of the proteins by Coomassie staining was inconclusive as to

autocleavage (Figure 8). Only the PN protein sample showed induced protein in the size ranges possible for SEA monomers. PC showed an induced protein band of 28 kDa, twice the expected size. This could be due to two copies of the PLN SEA being present in the expression plasmid. Western blot confirmed the PLN SEA data from the Coomassie staining (Figure 10). The PLN SEA band electrophoresed at a weight equal to that of the whole uncleaved recombinant protein. Models of the autocleavage regions of MUC1 and PLN indicate that the reason the PLN SEA does not autocleave may be attributed to two properties: the sequence and the fold. The analysis in chapter 2 shows the PLN SEA has a phenylalanine instead of a serine so there is no hydroxyl group is available to perform the N to O acyl rearrangement needed for the MUC1 SEA to autocleave (Figure 6) (Levitin et *al.*, 2005). The PLN SEA autocleavage model shows a longer and perhaps more flexible loop in between the two strands (Figure 11). This loop may be necessary for the packing of bulky amino acids like the tryptophan and phenylalanine in the PLN autocleavage motif and may also contribute to a non-cleaving behavior.

The predictions made in table 4 are based on sequence alignments and fold modeling. An alignment of SEA modules that autocleave show a 100% conservation of a phenylalanine, glycine, and a serine. The glycine and serine are those of the autocleavage motif. Fold modeling showed the identified phenylalanine next to the sequence that cleaves. The phenylalanine may apply pressure to the turn between the two β strands aiding in the ester formation and the resulting cleavage.

4.3. Binding proteins and their sub-cellular localization

The Y2H screen identified many proteins capable of binding the PLN SEA module. The most likely binding partners include AFAP1, EXOSC9, GABARAP, and FBXO28.

AFAP1 is a 730 amino acid, approximately 81 kDa protein that is a substrate for SRC and PKC (Flynn et *al.*, 1993; Qian et *al.*, 1998, 2002). It may play a role as an adaptor protein linking Src family members and other signaling proteins to actin filaments (Qian et *al.*, 2000; Baisden et *al.*, 2001). Phosphorylation sites that have been found at residues Y353, Y491, Y501, and S668. Multiple transcript variants encoding different isoforms have been found. Literature indicates it may play a role in regulating cell-matrix adhesions and cell migration during the development and progression of some cancers (Dorfleutner et *al.*, 2007; Walker et *al.*, 2007; Zhang et *al.*, 2007).

EXOSC9 is also called 75kD polymyositis/scleroderma autoantigen 1 or exosome complex exonuclease RPR45p6. It is a 439 amino acid protein that weighs approximately 49kD. To present, we know that it is a component of the exosome 3' → 5' exoribonuclease complex that degrades unstable mRNAs (Mukherjee et *al.*, 2002; Liu et *al.*, 2006; van Dijk et *al.*, 2007). EXOSC9 is subcellularly located in the cytoplasm, nucleus, and nucleolus. pSORT noted KKXX-like motif in the C-terminus (KRAA), which could function as an ER membrane

retention signal. pSORT also predicted subcellular localization primarily in the cytoplasm and nucleus, but a small portion was also predicted to be vacuolar. EXOSC9 autoantibodies are markers for autoimmune diseases, including scleroderma, systemic lupus erythematosus and polymyositis (Allmang et *al.*, 1999; Brouwer et *al.*, 2002; Raijmakers et *al.*, 2003).

GABARAP is 117 amino acids long and weighs approximately14 kDa. pSORT predicts GABARAP to localize in the cytoplasm, nucleus, and cytoskeleton. GABARAP binds neurotransmitter receptors by mediating interaction with the cytoskeleton (Wang et *al.*, 1999; Nymann-Andersen et *al.*, 2002). Literature also suggests localization in the endomembrane system and a function in intracellular membrane trafficking and autophagy (Kittler et *al.*, 2001; Kabeya et *al.*, 2004; Nakamura et *al.*, 2008; Kirkin et *al.*, 2009). GABARAP has been shown to have similar localization to calnexin in the ER and Sec23 at the ER exit site (Nakamura et al., 2008). A role in tumorigenesis via its intracellular trafficking function in invasive ductal carcinomas, invasive lobular carcinomas, adenomas and thyroid cancers also has been suggested (Klebig et *al.*, 2005; Roberts et *al.*, 2009).

FBXO28 forms Skp, Cullin, F-box containing complex (SCF complexes) to act as protein-ubiquitin ligases. This protein is 368 amino acids long and weighs approximately 41kDa, but much less is published on this protein. The protein degradation function of the SCF complexe makes FBXO28 less of a likely binding partner for the PLN SEA module.

GABARAP is interesting because PLN is processed through the secretory pathway and GABARAP is shown to be involved with the pathway through vesicle formation and binding. The other three Y2H screen high confidence hits have cellular topological conflicts that point to the possibility of experimental artifacts related to the way the yeast two hybrid system works.

Immunostaining was only performed on GABARAP of the potential binding proteins identified by Y2H. The individual ICC staining matched the expected patterns (Figure 17). PLN stained as punctate foci, similar to patterns expected of vesicles in the secretory pathway. GABARAP stained like vesicles as well but also stained diffusely throughout the cell. This matches information in the literature stating GABARAP is a cytosolic protein associated with vesicles (Wang et al., 2000; Kabeya et al., 2004; Nakamura et al., 2008). While GABARAP is associated with vesicles, it may never be found on the inner leaflet of the endomembrane system and thus not have access to PLN during trafficking. Co-localization by ICC staining was performed to study this. Dual staining of PLN and GABARAP in WiDr cells was inconclusive (Figure 18). The staining shows the two proteins are in the same regions of the cell but the resolution is not great enough to show protein-protein interactions. Refinement of standard ICC methods or imaging on a confocal microscopy or electron microscopy might provide intracellular localization information. Additionally protein complex immunoprecipitation could separately show protein binding.

4.4. Future Work

4.4.1. Structural analysis

In this work recombinant proteins were used to evaluate autocleavage of the PLN SEA module and to verify the autocleavage of the MUC1 SEA previously shown in the literature. Further analysis of the recombinant protein will help to support the finding that the PLN SEA module does not autocleave and ensure that is not artifact of the method of making and collecting the protein. Secondary structural analysis can be performed with circular dichroism. This will exclude protein misfolding as the reason for the PLN SEA module not autocleaving. Additionally tertiary structure analysis through the Protein Production Group (PPG) at MD Anderson Cancer Center can contribute more knowledge of the PLN SEA. The PPG is a standalone entity within the Center for Biomolecular Structure and Function that studies 3D protein structure by X-ray crystallography, NMR, or biophysical assay, only requiring an expression plasmid for submission.

4.4.2. Immunohistochemistry

Thus far co-localization studies have not been definitive. Further optimization of the staining protocol is needed to confirm co-localization of PLN and GABARAP. Confocal microscopy could be used to increase resolution of imaging and can slices through the cell to confirm that the molecules are present in the same plane. To complement standard co-localization a commercially available protein

binding assay called Duolink® proximity assay can be used in parallel. Duolink® PLA® technology uses oligonucleotide labeled secondary antibodies to detect protein interactions within a 40nm distance of one another.

4.4.3. Knockdown of PLN SEA module binding proteins

Once PLN SEA binding proteins are confirmed, their contribution to PLN trafficking and secretion can be studied. siRNA knockdown of candidate proteins should show whether these proteins are necessary for secretion or act to control the time spent in the secretory pathway. A change in the amount of PLN secreted into the media can be measured after siRNA treatment to determine if the protein is required for successful secretion. Additionally a pulse-chase experiment could be used to determine if binding proteins control the time spent in the pathway.

4.4.4. Contribution of residues involved in binding

Lastly, the role of the residues identified in chapter 2 as possibly being involved in binding should be examined. Each residue could be changed individually to a hydrophilic amino acid and expressed within a small portion of the protein containing the SEA in mammalian cells. The purpose would be to see if altering their hydrophobicity and possible the ability to bind other proteins would change the path or time for normal protein maturation and presentation.

I would also like to use mutational methods to change the autocleavage motif in the PLN SEA to try to induce autocleavage. Initially I would replace the tryptophan in the GWVFV PLN sequence with a serine. The introduction of the serine at this position should be enough to cause autocleavage if the mechanism alone described in chapter one if sufficient to cause autoproteolysis. The phenylalanine could also be changed in combination with the tryptophan to restore the exact autocleavage motif from the MUC1. I would like to see if changing the peptide sequence would be enough or if the longer loop between the β strands would still block autocleavage. If autocleavage still does not occur then a phenylalanine could replace the tyrosine that sits where phenylalanine was shown to possibly play a role in autocleavage in chapter two. If autocleavage can be induced in the PLN SEA, I would like to the expression a small portion of PLN with the modified SEA to see if autocleavage alters the transfer of the protein from the ER to the Golgi or slows down the transfer. It would also be interesting to see if autocleavage changes the composition of the carbohydrates added just before the SEA.

References

Abe, J., Fukuzawa, T., and Hirose, S. (2002a). Cleavage of Ig-Hepta at a “SEA” module and at a conserved G protein-coupled receptor proteolytic site. J. Biol. Chem. *277*, 23391–23398.

Abe, J., Fukuzawa, T., and Hirose, S. (2002b). Cleavage of Ig-Hepta at a “SEA” module and at a conserved G protein-coupled receptor proteolytic site. J. Biol. Chem. *277*, 23391–23398.

Aberfeld, D.C., Hinterbuchner, L.P., and Schneider, M. (1965). Myotonia, dwarfism, diffuse bone disease and unusual ocular and facial abnormalities (a new syndrome). Brain *88*, 313–322.

Allmang, C., Petfalski, E., Podtelejnikov, A., Mann, M., Tollervey, D., and Mitchell, P. (1999). The yeast exosome and human PM–Scl are related complexes of $3' \rightarrow 5'$ exonucleases. Genes Dev. *13*, 2148–2158.

Baisden, J.M., Gatesman, A.S., Cherezova, L., Jiang, B.H., and Flynn, D.C. (2001). The intrinsic ability of AFAP-110 to alter actin filament integrity is linked with its ability to also activate cellular tyrosine kinases. Oncogene *20*, 6607–6616.

Bandah-Rozenfeld, D., Collin, R.W.J., Banin, E., van den Born, L.I., Coene, K.L.M., Siemiatkowska, A.M., Zelinger, L., Khan, M.I., Lefeber, D.J., Erdinest, I., et al. (2010). Mutations in IMPG2, encoding interphotoreceptor matrix proteoglycan 2, cause autosomal-recessive retinitis pigmentosa. Am. J. Hum. Genet. *87*, 199–208.

Bartel, P.L., Chien, C.T., Sternglanz, R., and Fields, S. (1993). Using the two-hybrid system to detect protein-protein interactions. In Cellular Interactions in Development: A Practical Approach, (Oxford: Oxford University Press), pp. 153–179.

Bengtsson, E., Mörgelin, M., Sasaki, T., Timpl, R., Heinegård, D., and Aspberg, A. (2002). The leucine-rich repeat protein PRELP binds perlecan and collagens and may function as a basement membrane anchor. J. Biol. Chem. *277*, 15061–15068.

Bix, G., Castello, R., Burrows, M., Zoeller, J.J., Weech, M., Iozzo, R.A., Cardi, C., Thakur, M.L., Barker, C.A., Camphausen, K., et al. (2006). Endorepellin in vivo: targeting the tumor vasculature and retarding cancer growth and metabolism. J. Natl. Cancer Inst. *98*, 1634–1646.

Bix, G., Fu, J., Gonzalez, E.M., Macro, L., Barker, A., Campbell, S., Zutter, M.M., Santoro, S.A., Kim, J.K., Höök, M., et al. (2004). Endorepellin causes endothelial cell disassembly of actin cytoskeleton and focal adhesions through alpha2beta1 integrin. J. Cell Biol. *166*, 97–109.

Bork, P., and Patthy, L. (1995). The SEA module: a new extracellular domain associated with O-glycosylation. Protein Sci. *4*, 1421–1425.

Brouwer, R., Vree Egberts, W.T.M., Hengstman, G.J.D., Raijmakers, R., van Engelen, B.G.M., Seelig, H.P., Renz, M., Mierau, R., Genth, E., Pruijn, G.J.M., et al. (2002). Autoantibodies directed to novel components of the PM/Scl complex, the human exosome. Arthritis Res. *4*, 134–138.

Brown, A.J., Alicknavitch, M., D'Souza, S.S., Daikoku, T., Kirn-Safran, C.B., Marchetti, D., Carson, D.D., and Farach-Carson, M.C. (2008). Heparanase Expression and Activity Influences Chondrogenic and Osteogenic Processes During Endochondral Bone Formation. Bone *43*, 689–699.

Chang, Z., Meyer, K., Rapraeger, A.C., and Friedl, A. (2000). Differential ability of heparan sulfate proteoglycans to assemble the fibroblast growth factor receptor complex in situ. Faseb J. *14*, 137–144.

Clayton, A., Thomas, J., Thomas, G.J., Davies, M., and Steadman, R. (2001). Cell surface heparan sulfate proteoglycans control the response of renal interstitial fibroblasts to fibroblast growth factor-2. Kidney Int. *59*, 2084–2094.

Costell, M., Gustafsson, E., Aszódi, A., Mörgelin, M., Bloch, W., Hunziker, E., Addicks, K., Timpl, R., and Fässler, R. (1999). Perlecan maintains the integrity of cartilage and some basement membranes. J. Cell Biol. *147*, 1109–1122.

Costell, M., Mann, K., Yamada, Y., and Timpl, R. (1997). Characterization of recombinant perlecan domain I and its substitution by glycosaminoglycans and oligosaccharides. Eur. J. Biochem. *243*, 115–121.

van Dijk, E.L., Schilders, G., and Pruijn, G.J.M. (2007). Human cell growth requires a functional cytoplasmic exosome, which is involved in various mRNA decay pathways. RNA *13*, 1027–1035.

Dolan, M., Horchar, T., Rigatti, B., and Hassell, J.R. (1997). Identification of sites in domain I of perlecan that regulate heparan sulfate synthesis. J. Biol. Chem. *272*, 4316–4322.

Dorfleutner, A., Stehlik, C., Zhang, J., Gallick, G.E., and Flynn, D.C. (2007). AFAP-110 is required for actin stress fiber formation and cell adhesion in MDA-MB-231 breast cancer cells. J. Cell. Physiol. *213*, 740–749.

Ettner, N., Göhring, W., Sasaki, T., Mann, K., and Timpl, R. (1998). The N-terminal globular domain of the laminin alpha1 chain binds to alpha1beta1 and alpha2beta1 integrins and to the heparan sulfate-containing domains of perlecan. FEBS Lett. *430*, 217–221.

Fjeldstad, K., and Kolset, S.O. (2005). Decreasing the metastatic potential in cancers--targeting the heparan sulfate proteoglycans. Curr Drug Targets *6*, 665–682.

Flynn, D.C., Leu, T.H., Reynolds, A.B., and Parsons, J.T. (1993). Identification and sequence analysis of cDNAs encoding a 110-kilodalton actin filament-associated pp60src substrate. Mol Cell Biol *13*, 7892–7900.

Formstecher, E., Aresta, S., Collura, V., Hamburger, A., Meil, A., Trehin, A., Reverdy, C., Betin, V., Maire, S., Brun, C., et al. (2005). Protein interaction mapping: a Drosophila case study. Genome Res. *15*, 376–384.

French, M.M., Smith, S.E., Akanbi, K., Sanford, T., Hecht, J., Farach-Carson, M.C., and Carson, D.D. (1999). Expression of the heparan sulfate proteoglycan, perlecan, during mouse embryogenesis and perlecan chondrogenic activity in vitro. J. Cell Biol. *145*, 1103–1115.

Friedrich, M.V., Göhring, W., Mörgelin, M., Brancaccio, A., David, G., and Timpl, R. (1999). Structural basis of glycosaminoglycan modification and of heterotypic interactions of perlecan domain V. J. Mol. Biol. *294*, 259–270.

Fromont-Racine, M., Rain, J.-C., and Legrain, P. (1997). Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. Nature Genetics *16*, 277–282.

Ghiselli, G., Eichstetter, I., and Iozzo, R.V. (2001). A role for the perlecan protein core in the activation of the keratinocyte growth factor receptor. Biochem. J. *359*, 153–163.

Giedion, A. (1997). Heterogeneity in Schwartz-Jampel chondrodystrophic myotonia. Pediatr Radiol *27*, 454.

Goyal, A., Pal, N., Concannon, M., Paul, M., Doran, M., Poluzzi, C., Sekiguchi, K., Whitelock, J.M., Neill, T., and Iozzo, R.V. (2011). Endorepellin, the angiostatic module of perlecan, interacts with both the α2β1 integrin and vascular endothelial growth factor receptor 2 (VEGFR2): a dual receptor antagonism. J. Biol. Chem. *286*, 25947–25962.

Hassell, J.R., Robey, P.G., Barrach, H.J., Wilczek, J., Rennard, S.I., and Martin, G.R. (1980). Isolation of a heparan sulfate-containing proteoglycan from basement membrane. Proc. Natl. Acad. Sci. U.S.A. *77*, 4494–4498.

Ho, S.B., Luu, Y., Shekels, L.L., Batra, S.K., Kandarian, B., Evans, D.B., Zaworski, P.G., Wolfe, C.L., and Heinrikson, R.L. (2010). Activity of recombinant cysteine-rich domain proteins derived from the membrane-bound MUC17/Muc3 family mucins. Biochim. Biophys. Acta *1800*, 629–638.

Hopf, M., Göhring, W., Kohfeldt, E., Yamada, Y., and Timpl, R. (1999). Recombinant domain IV of perlecan binds to nidogens, laminin-nidogen complex, fibronectin, fibulin-2 and heparin. Eur. J. Biochem. *259*, 917–925.

Hopf, M., Göhring, W., Mann, K., and Timpl, R. (2001). Mapping of binding sites for nidogens, fibulin-2, fibronectin and heparin to different IG modules of perlecan. J. Mol. Biol. *311*, 529–541.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J Mol Graph *14*, 33–38, 27–28.

Iozzo, R.V. (1984). Biosynthesis of heparan sulfate proteoglycan by human colon carcinoma cells and its localization at the cell surface. J. Cell Biol. *99*, 403–417.

Kabeya, Y., Mizushima, N., Yamamoto, A., Oshitani-Okamoto, S., Ohsumi, Y., and Yoshimori, T. (2004). LC3, GABARAP and GATE16 localize to autophagosomal membrane depending on form-II formation. J Cell Sci *117*, 2805–2812.

Kelley, L.A., and Sternberg, M.J.E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc *4*, 363–371.

Khatri, I.A., Wang, R., and Forstner, J.F. (2003). SEA (sea-urchin sperm protein, enterokinase and agrin)-module cleavage, association of fragments and membrane targeting of rat intestinal mucin Muc3. Biochem. J. *372*, 263–270.

Kirkin, V., Lamark, T., Sou, Y.-S., Bjørkøy, G., Nunn, J.L., Bruun, J.-A., Shvets, E., McEwan, D.G., Clausen, T.H., Wild, P., et al. (2009). A role for NBR1 in autophagosomal degradation of ubiquitinated substrates. Mol. Cell *33*, 505–516.

Kittler, J.T., Rostaing, P., Schiavo, G., Fritschy, J.M., Olsen, R., Triller, A., and Moss, S.J. (2001). The subcellular distribution of GABARAP and its ability to interact with NSF suggest a role for this protein in the intracellular transport of GABA(A) receptors. Mol. Cell. Neurosci. *18*, 13–25.

Klebig, C., Seitz, S., Arnold, W., Deutschmann, N., Pacyna-Gengelbach, M., Scherneck, S., and Petersen, I. (2005). Characterization of {gamma}-aminobutyric acid type A receptor-associated protein, a novel tumor suppressor, showing reduced expression in breast cancer. Cancer Res. *65*, 394–400.

Kohfeldt, E., Sasaki, T., Göhring, W., and Timpl, R. (1998). Nidogen-2: a new basement membrane protein with diverse binding properties. J. Mol. Biol. *282*, 99–109.

Kokenyesi, R., and Silbert, J.E. (1995). Formation of heparan sulfate or chondroitin/dermatan sulfate on recombinant domain I of mouse perlecan

expressed in Chinese hamster ovary cells. Biochem. Biophys. Res. Commun. *211*, 262–267.

Kranz, J.K., Lu, J., and Hall, K.B. (1996). Contribution of the tyrosines to the structure and function of the human U1A N-terminal RNA binding domain. Protein Science *5*, 1567–1583.

Kvansakul, M., Hopf, M., Ries, A., Timpl, R., and Hohenester, E. (2001). Structural basis for the high-affinity interaction of nidogen-1 with immunoglobulin-like domain 3 of perlecan. Embo J *20*, 5342–5346.

Levitin, F., Stern, O., Weiss, M., Gil-Henn, C., Ziv, R., Prokocimer, Z., Smorodinsky, N.I., Rubinstein, D.B., and Wreschner, D.H. (2005). The MUC1 SEA module is a self-cleaving domain. J. Biol. Chem. *280*, 33374–33386.

Lillehoj, E.P., Han, F., and Kim, K.C. (2003). Mutagenesis of a Gly-Ser cleavage site in MUC1 inhibits ectodomain shedding. Biochem. Biophys. Res. Commun. *307*, 743–749.

Liu, Q., Greimann, J.C., and Lima, C.D. (2006). Reconstitution, activities, and structure of the eukaryotic RNA exosome. Cell *127*, 1223–1237.

Macao, B., Johansson, D.G.A., Hansson, G.C., and Härd, T. (2006). Autoproteolysis coupled to protein folding in the SEA domain of the membrane-bound MUC1 mucin. Nat. Struct. Mol. Biol. *13*, 71–76.

Maeda, T., Inoue, M., Koshiba, S., Yabuki, T., Aoki, M., Nunokawa, E., Seki, E., Matsuda, T., Motoda, Y., Kobayashi, A., et al. (2004). Solution structure of the SEA domain from the murine homologue of ovarian cancer antigen CA125 (MUC16). J. Biol. Chem. *279*, 13174–13182.

Melrose, J., Roughley, P., Knox, S., Smith, S., Lord, M., and Whitelock, J. (2006). The Structure, Location, and Function of Perlecan, a Prominent Pericellular Proteoglycan of Fetal, Postnatal, and Mature Hyaline Cartilages. J. Biol. Chem. *281*, 36905–36914.

Miettinen, A., Stow, J.L., Mentone, S., and Farquhar, M.G. (1986). Antibodies to Basement Membrane Heparan Sulfate Proteoglycans Bind to the Laminae Rarae of the Glomerular Basement Membrane (GBM) and Induce Subepithelial GBM Thickening. J Exp Med *163*, 1064–1084.

Miosge, N., Simniok, T., Sprysch, P., and Herken, R. (2003). The collagen type XVIII endostatin domain is co-localized with perlecan in basement membranes in vivo. J. Histochem. Cytochem. *51*, 285–296.

Mongiat, M., Sweeney, S.M., San Antonio, J.D., Fu, J., and Iozzo, R.V. (2003). Endorepellin, a novel inhibitor of angiogenesis derived from the C terminus of perlecan. J. Biol. Chem. *278*, 4238–4249.

Mongiat, M., Taylor, K., Otto, J., Aho, S., Uitto, J., Whitelock, J.M., and Iozzo, R.V. (2000). The protein core of the proteoglycan perlecan binds specifically to fibroblast growth factor-7. J. Biol. Chem. *275*, 7095–7100.

Morita, H., Yoshimura, A., Inui, K., Ideura, T., Watanabe, H., Wang, L., Soininen, R., and Tryggvason, K. (2005). Heparan sulfate of perlecan is involved in glomerular filtration. J. Am. Soc. Nephrol. *16*, 1703–1710.

Mukherjee, D., Gao, M., O'Connor, J.P., Raijmakers, R., Pruijn, G., Lutz, C.S., and Wilusz, J. (2002). The mammalian exosome mediates the efficient degradation of mRNAs that contain AU-rich elements. The EMBO Journal *21*, 165–174.

Murdoch, A.D., Dodge, G.R., Cohen, I., Tuan, R.S., and Iozzo, R.V. (1992). Primary structure of the human heparan sulfate proteoglycan from basement membrane (HSPG2/perlecan). A chimeric molecule with multiple domains homologous to the low density lipoprotein receptor, laminin, neural cell adhesion molecules, and epidermal growth factor. J. Biol. Chem. *267*, 8544–8557.

Nakamura, T., Hayashi, T., Nasu-Nishimura, Y., Sakaue, F., Morishita, Y., Okabe, T., Ohwada, S., Matsuura, K., and Akiyama, T. (2008). PX-RICS mediates ER-to-Golgi transport of the N-cadherin/beta-catenin complex. Genes Dev. *22*, 1244–1256.

Noonan, D.M., and Hassell, J.R. (1993). Perlecan, the large low-density proteoglycan of basement membranes: structure and variant forms. Kidney Int. *43*, 53–60.

Nymann-Andersen, J., Wang, H., Chen, L., Kittler, J.T., Moss, S.J., and Olsen, R.W. (2002). Subunit specificity and interaction domain between GABA(A) receptor-associated protein (GABARAP) and GABA(A) receptors. J. Neurochem. *80*, 815–823.

Palmai-Pallag, T., Khodabukus, N., Kinarsky, L., Leir, S.-H., Sherman, S., Hollingsworth, M.A., and Harris, A. (2005). The role of the SEA (sea urchin sperm protein, enterokinase and agrin) module in cleavage of membrane-tethered mucins. Febs J. *272*, 2901–2911.

Parry, S., Silverman, H.S., McDermott, K., Willis, A., Hollingsworth, M.A., and Harris, A. (2001). Identification of MUC1 proteolytic cleavage sites in vivo. Biochem. Biophys. Res. Commun. *283*, 715–720.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605–1612.

Qian, Y., Baisden, J.M., Cherezova, L., Summy, J.M., Guappone-Koay, A., Shi, X., Mast, T., Pustula, J., Zot, H.G., Mazloum, N., et al. (2002). PC phosphorylation increases the ability of AFAP-110 to cross-link actin filaments. Mol. Biol. Cell *13*, 2311–2322.

Qian, Y., Baisden, J.M., Westin, E.H., Guappone, A.C., Koay, T.C., and Flynn, D.C. (1998). Src can regulate carboxy terminal interactions with AFAP-110, which influence self-association, cell localization and actin filament integrity. Oncogene *16*, 2185–2195.

Qian, Y., Baisden, J.M., Zot, H.G., Van Winkle, W.B., and Flynn, D.C. (2000). The carboxy terminus of AFAP-110 modulates direct interactions with actin filaments and regulates its ability to alter actin filament integrity and induce lamellipodia formation. Exp. Cell Res. *255*, 102–113.

Raijmakers, R., Egberts, W.V., van Venrooij, W.J., and Pruijn, G.J.M. (2003). The Association of the Human PM/Scl-75 Autoantigen with the Exosome Is Dependent on a Newly Identified N Terminus. J. Biol. Chem. *278*, 30698–30704.

Rain, J.C., Selig, L., De Reuse, H., Battaglia, V., Reverdy, C., Simon, S., Lenzen, G., Petel, F., Wojcik, J., Schächter, V., et al. (2001). The protein-protein interaction map of Helicobacter pylori. Nature *409*, 211–215.

Roberts, S.S., Mendonça-Torres, M.C., Jensen, K., Francis, G.L., and Vasko, V. (2009). GABA receptor expression in benign and malignant thyroid tumors. Pathol. Oncol. Res. *15*, 645–650.

Rodgers, K.D., Sasaki, T., Aszodi, A., and Jacenko, O. (2007). Reduced perlecan in mice results in chondrodysplasia resembling Schwartz-Jampel syndrome. Hum. Mol. Genet. *16*, 515–528.

Salmivirta, K., Talts, J.F., Olsson, M., Sasaki, T., Timpl, R., and Ekblom, P. (2002). Binding of mouse nidogen-2 to basement membrane components and cells and its expression in embryonic and adult tissues suggest complementary functions of the two nidogens. Exp. Cell Res. *279*, 188–201.

Sasaki, T., Costell, M., Mann, K., and Timpl, R. (1998). Inhibition of glycosaminoglycan modification of perlecan domain I by site-directed mutagenesis changes protease sensitivity and laminin-1 binding activity. FEBS Letters *435*, 169–172.

Sasaki, T., Göhring, W., Pan, T.C., Chu, M.L., and Timpl, R. (1995). Binding of mouse and human fibulin-2 to extracellular matrix ligands. J. Mol. Biol. *254*, 892–899.

Sasse, P., Malan, D., Fleischmann, M., Roell, W., Gustafsson, E., Bostani, T., Fan, Y., Kolbe, T., Breitbach, M., Addicks, K., et al. (2008). Perlecan is critical for heart stability. Cardiovasc. Res. *80*, 435–444.

Schofield, K.P., Gallagher, J.T., and David, G. (1999). Expression of proteoglycan core proteins in human bone marrow stroma. Biochem. J. *343 Pt 3*, 663–668.

Schweitzer, K.M., Vicart, P., Delouis, C., Paulin, D., Dräger, A.M., Langenhuijsen, M.M., and Weksler, B.B. (1997). Characterization of a newly established human bone marrow endothelial cell line: distinct adhesive properties for hematopoietic progenitors compared with human umbilical vein endothelial cells. Lab. Invest. *76*, 25–36.

Smeland, S., Kolset, S.O., Lyon, M., Norum, K.R., and Blomhoff, R. (1997). Binding of perlecan to transthyretin in vitro. Biochem. J. *326 (Pt 3)*, 829–836.

Spranger, J., Hall, B.D., Häne, B., Srivastava, A., and Stevenson, R.E. (2000). Spectrum of Schwartz-Jampel syndrome includes micromelic chondrodysplasia, kyphomelic dysplasia, and Burton disease. Am. J. Med. Genet. *94*, 287–295.

Tapanadechopone, P., Hassell, J.R., Rigatti, B., and Couchman, J.R. (1999). Localization of glycosaminoglycan substitution sites on domain V of mouse perlecan. Biochem. Biophys. Res. Commun. *265*, 680–690.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673–4680.

Tu, H., Sasaki, T., Snellman, A., Göhring, W., Pirilä, P., Timpl, R., and Pihlajaniemi, T. (2002). The type XIII collagen ectodomain is a 150-nm rod and capable of binding to fibronectin, nidogen-2, perlecan, and heparin. J. Biol. Chem. *277*, 23092–23099.

Vojtek, A.B., and Hollenberg, S.M. (1995). Ras-Raf interaction: two-hybrid analysis. Meth. Enzymol. *255*, 331–342.

Walker, V.G., Ammer, A., Cao, Z., Clump, A.C., Jiang, B.-H., Kelley, L.C., Weed, S.A., Zot, H., and Flynn, D.C. (2007). PI3K activation is required for PMA-directed activation of cSrc by AFAP-110. Am J Physiol Cell Physiol *293*, C119–C132.

Wang, H., Bedford, F.K., Brandon, N.J., Moss, S.J., and Olsen, R.W. (1999). GABA(A)-receptor-associated protein links GABA(A) receptors and the cytoskeleton. Nature *397*, 69–72.

Wang, H., Olsen, R.W., Wang, H., and Olsen, R.W. (2000). Binding of the GABAA Receptor-Associated Protein (GABARAP) to Microtubules and Microfilaments Suggests Involvement of the Cytoskeleton in GABARAPGABAA Receptor Interaction, Binding of the GABAA Receptor-Associated Protein (GABARAP) to Microtubules and Microfilaments Suggests Involvement of the Cytoskeleton in GABARAPGABAA

Receptor Interaction. Journal of Neurochemistry, Journal of Neurochemistry *75, 75*, 644, 644–655, 655.

Whitelock, J.M., Graham, L.D., Melrose, J., Murdoch, A.D., Iozzo, R.V., and Anne Underwood, P. (1999). Human perlecan immunopurified from different endothelial cell sources has different adhesive properties for vascular cells. Matrix Biology *18*, 163–178.

Wojcik, J., Boneca, I.G., and Legrain, P. (2002). Prediction, assessment and validation of protein interaction maps in bacteria. J. Mol. Biol. *323*, 763–770.

Wreschner, D.H., McGuckin, M.A., Williams, S.J., Baruch, A., Yoeli, M., Ziv, R., Okun, L., Zaretsky, J., Smorodinsky, N., Keydar, I., et al. (2002). Generation of ligand-receptor alliances by "SEA" module-mediated cleavage of membrane-associated mucin proteins. Protein Sci. *11*, 698–706.

Zhang, J., Park, S.I., Artime, M.C., Summy, J.M., Shah, A.N., Bomser, J.A., Dorfleutner, A., Flynn, D.C., and Gallick, G.E. (2007). AFAP-110 is overexpressed in prostate cancer and contributes to tumorigenic growth by regulating focal contacts. J. Clin. Invest. *117*, 2962–2973.

Zilberfarb, V., Piétri-Rouxel, F., Jockers, R., Krief, S., Delouis, C., Issad, T., and Strosberg, A.D. (1997). Human immortalized brown adipocytes express functional beta3-adrenoceptor coupled to lipolysis. J. Cell. Sci. *110 (Pt 7)*, 801–807.

Online Mendelian Inheritance in Man, OMIM®. Johns Hopkins University, Baltimore, MD. MIM Number: 142461: 01/28/2011: . World Wide Web URL: http://omim.org/entry/142461